A Generalized Partial Credit FACETS Model for Investigating Order Effects in Self-Report Personality Data

A Dissertation

Presented to

The Academic Faculty

by

Heather Hayes

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy in the

School of Psychology

Georgia Institute of Technology

August 2012

A Generalized Partial Credit FACETS Model for Investigating Order Effects in Self-Report Personality Data

Approved by:

Dr. Susan E. Embretson, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Lawrence James
School of Psychology
Georgia Institute of Technology

Dr. Jack Feldman
School of Psychology
Georgia Institute of Technology

Dr. Davood Tofighi
School of Psychology
Georgia Institute of Technology

Dr. Charles Parsons
College of Management
Georgia Institute of Technology

Date Approved: June 27, 2012

ACKNOWLEDGEMENTS

First and foremost, I want to thank my parents, Arthur and Kathleen Hayes, for giving me an amazing, Griswold-esque childhood. They were always supportive of anything I did – especially music and science – and attended every single event in which I was involved or performed. My dad's stellar career, tenacity, and perseverance have always inspired me to pursue a field involving higher education. My mother has always been nurturing, selfless, and giving while providing wise advice. My sister, Kristin, who also has a doctorate, continues to inspire me with her accomplishments and her lifestyle, and she is my best friend in the world. I want to thank my son, Paul, for keeping me sane when times were tough, providing unconditional love and serving as the greatest motivation for me to work hard and make him proud, someday. He is the love of my life. Paul's father, Adam, is a wonderful father and, in the early years of my Ph.D. work, was supportive in ways that I cannot even describe. I will always love him. I also want to thank his mom, Deborah, and her husband, Ross, for making the best wine ever and being such a wonderful second family to me – as well as close friends – all of these years. I have many wonderful friends – Jess, Mandy, Holly, Jamie, Noelle, and Aimee – who I've known for many years and who have also inspired me to be successful because they, themselves, are all well-educated, career-oriented, successful individuals that I highly admire and look up to. I want to thank Michael Francisco for being here for me when it really counted – the end. I love him dearly and always will. Nobody will understand or know me as well as him with the exception, perhaps, of my parents and sister. Finally, I want to thank Susan Embretson for being an amazing advisor and teacher. I came to this school, seeing her as a celebrity, and I still do. I am eternally honored to call her my mentor and advisor. All of the psychology professors at Georgia Tech are excellent, so I want to thank them all. But, to be specific: Larry James for being the most fun to work with, research-wise, and for the extensive knowledge I gained from him regarding personality theory; Jack Feldman for being

generally supportive and enthusiastic about my work, while providing direction as well; Charles Parsons for being on my committee, being genuinely nice and approachable, and letting me teach statistics in the College of Management; Davood Tofighi for being on my committee and providing advice with respect to Bayesian analysis; and finally, Jim Roberts for helping me throughout my years here, making me feel like I had a home when I had none, being perhaps the best professor I've ever had, and aiding me during the first half of my dissertation work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Summary

Despite its convenience, the process of self-report in personality testing can be impacted by a variety of cognitive and perceptual biases. One bias that violates local independence, a core criterion of modern test theory, is the order effect. In this bias, characteristics of an item response are impacted not only by the content of the current item but also the accumulated exposure to previous, similar-content items. This bias is manifested as increasingly stable item responses for items that appear later in a test. Previous investigations of this effect have been rooted in classical test theory (CTT) and have consistently found that item reliabilities, or corrected item-total score correlations, increase with the item's serial position in the test. The purpose of the current study was to more rigorously examine order effects via item response theory (IRT). To this end, the FACETS modeling approach (Linacre, 1989) was combined with the Generalized Partial Credit model (GPCM; Muraki, 1992) to produce a new model, the Generalized Partial Credit FACETS model (GPCFM). Serial position of an item serves as a facet that contributes to the item response, not only via its impact on an item's location on the latent trait continuum, but also its discrimination. Thus, the GPCFM differs from previous generalizations of the FACETS model (Wang & Liu, 2007) in that the item discrimination parameter is modified to include a serial position effect. This parameter is important because it reflects the extent to which the purported underlying trait is represented in an item score. Two sets of analyses were conducted. First, a simulation study demonstrated effective parameter recovery, though measurements of error were impacted by sample size for all parameters, test length for trait level estimates, and the size of the order effect for trait level estimates, and an interaction between sample size and test length for item discrimination. Secondly, with respect to real self-report personality data, the GPCFM demonstrated good fit as well as superior fit relative to competing, nested models while also identifying order effects in some traits, particularly Neuroticism, Openness, and Agreeableness.

CHAPTER I

INTRODUCTION

Self-report personality scores are meant to serve as proxies for observing patterns of behavior over time, thereby representing an individual's average standings on theoretically stable, latent personality traits. However, when more closely examined, the *process* by which a respondent rates him or herself on a given trait item may, in fact, reflect a phenomenon that potentially limits the extent to which the item score does, indeed, correspond with a stable pattern of behavior. Specifically, the response process is, arguably, reflective of social-cognitive factors such as self-perception and self-concept (Cervone, Shadel, & Jencius, 2001; Markus, 1977), unconscious processes and perceptual bias (Greenwald & Banaji, 1995), degree of self-awareness (Levine, Huff, Wagner, & Sweeney, 1998; Nasby, 1989), one's mood at that time (Harris & Lucia, 2003; Kihlstrom, Eich, Sandbrand, & Tobias, 2000) and situational constraints (Menon & Yorkston, 2000; Mischel & Shoda, 1999; Schwarz, 1999; Tourangeau, 2000). For example, respondents are being measured on level of conscientiousness when applying for a job. With respect to situational constraints that increase motivation to present one's self in a desirable manner, a respondent is aware that high levels of conscientiousness are valued by potential employers and, thus, reports performing behaviors consistent with this trait conscientiousness (e.g., achievement motivation; dutifulness). Again, however, the respondent usually does not, in fact, behave in ways reflective of this trait. In contrast, in an unconstrained setting where responses are anonymous and perceptual biases may occur, the respondent may believe him or herself to be conscientious, selectively retrieving memories or examples when he or she was, indeed, conscientious when, he or she is, in fact, low on this trait. Even further, biases in self-awareness means that a respondent may not be familiar with the trait and struggles to recall examples of when he or she behaved conscientiously.

In essence, the mere process of measuring a latent trait – particularly personality - complicates the meaning behind a trait estimate. It is during this process in which certain perceptual and cognitive biases can emerge. For example, most individuals have high self-esteem and are optimistic about themselves (e.g., Greenwald & Banaji, 1995). They want to protect these positive self-concepts and will do so by managing their self-presentation in a manner that is socially desirable (e.g., Cramer, 1993; Nisbett & Ross, 1980; Nisbett & Wilson, 1977; Paulhus, 2002).

Given these issues, some research endeavors in personality measurement have focused on explicating the specific cognitive processes and strategies engaged during the self-report response process in order to determine their impact on the ultimate item response (Holtgraves, 2004; Holden, Kroner, & Popham, 1992; Fekken, Fekken & Holden, 1992). According to Tourangeau and Rasinski (1988), the response process consists of four stages. First, the respondent interprets the item stem and determines what type of behavior, characteristic, or trait is being assessed. Second, the respondent scans memories of his or her behavior for said behaviors that are characteristic of the trait and retrieves any relevant information. Third, the respondent compares retrieved memories, if any, to the item stem in order to determine the extent to which the item describes the respondent. This stage may involve integrating and consolidating multiple instances of the behavior or characteristic. In the final stage, a response is selected based on the degree to which the information in the item is perceived to match the respondent's retrieved memories or perceptions of the self. The latter stage is also likely to be affected by the response scale (e.g., yes or no; Likert scale).

Results of several studies have demonstrated that outcomes from the stages of the response process for any given item, or set of items, do indeed differ across respondents who vary in self-awareness, self-deceptive positivity (in which respondents honestly think highly of themselves regardless of reality; Paulhus, 2002) , and motivation to fake. The latter factor can be manifested

as either individual differences in motivation to fake or a situational context in which faking is an attractive strategy and thus can be experimentally manipulated (i.e., job selection). There is evidence to support changes in the response process – or particular aspects of it – not only across multiple test sessions (and differing contexts) but throughout a single test session (e.g., Holtgraves, 2004).

It has been argued that as the respondent answers more items in the survey, the response process becomes more streamlined because the respondent becomes more familiar with the item content. For example, suppose a respondent begins a test meant to measure the unidimensional trait of openness. As the respondent progresses through the test, one becomes increasingly aware of what trait is being measured due to repetitive exposure to overlapping, similar item characteristics. For example, the respondent typically requires less and less time to interpret the item stem and select a response. As will be more thoroughly explain in subsequent sections, cognitive theory can be used to explain this change in response process based on increased sensitivity of the respondent to schemas involving the trait being measured. In other words, the respondent more easily and quickly accesses this information (trait schemas) throughout the test and determines whether it fits the respondent's self-schema, resulting in quicker and more consistent responses as more items are encountered.

From a classical test theory perspective (CTT; Lord & Novick, 1968), the result of this change in response process is an increase in the correlation between an item score and the total test score as the item appears later in the test (Hamilton & Schuminsky, 1990; Knowles, 1988; Knowles & Byers, 1996; Ostrom, Betz, & Skowronski, 1992). This phenomenon, in which "measurement changes the measure" (Knowles, 1988) is interchangeably referred to as context effects or (serial) order effects. This phenomenon threatens the validity of self-report personality score interpretation, as briefly reviewed. However, as will be elaborated upon shortly, more sophisticated modeling techniques such as IRT have not been sufficiently applied to testing the

presence and size of this bias. This is the goal of the current study. Namely, a generalized facets IRT model will be constructed for graded responses in order to account for the impact of order on an item's ability to differentiate respondents with different trait levels as well as location of the item on the latent trait continuum.

Simulated data will be used to test for model efficacy and parameter recovery with various sample sizes and test lengths. Once the new model has been used to generate simulated data with known parameter values, if there are no substantial discrepancies between these true values and the estimated values, then there is evidence that, in theory, the model may be useful. Next, the model will be applied to the detection of context effects in a real data sample, followed by a comparison of fit for the new model to models that are nested within this new model, such as the GPCM (Muraki, 1992) and generalized FACETS model (Wang & Liu, 2007). Thus, the proposed study will potentially bolster previous findings on order effects in self-report by replicating them under more rigorous testing with a new IRT model developed to account for the impact of order on item discrimination and location. It is this context-based, order bias to which the focus of the current proposal turns.

CHAPTER II

CONTEXT EFFECTS IN LATENT TRAIT MEASUREMENT

The study of context effects in latent trait measurement is by no means a new topic of study. Research on this effect can be found in the literature from multiple construct domains, including personality and educational/achievement testing. In the latter, change in item parameter (difficulty and/or discrimination) over time is referred to as item parameter drift (IPD; Goldstein, 1983). Much of the research on stability of item parameters has focused on consistency across multiple test sessions (e.g., Bock, Muraki, & Pfeiffenberger, 1988; Goldstein, 1983; Mislevy, 1982; Sykes & Fitzpatrick, 1992). In general, these studies have found that IPD occurs as a function of practice effects, item content exposure and familiarity (Mislevy, 1982), changes in culture, education, and technology (Bock et al., 1988), as well as learning of a particular content area (Sykes & Fitzgerald, 1992). Specifically, item responses become more stable over time or across test sessions. Research has also shown that item parameters of an IRT model will vary across respondents, or groups of respondents, as a function of item order within a single test session (Leary & Dorans, 1985; Zwick, 1991). The order effect has been most dominant in reading comprehension tests, affecting not only item parameters for individual items but also for testlets or sets of items that accompany a single reading passage (Leary & Dorans, 1985).

In personality testing, concern initially emerged over the stability of test scores across multiple test sessions (Mischel, 1968). For example, scores were more reliable in later sessions, even if different items were used (Goldberg, 1978; Hayes, 1964). This phenomenon also occurred in single test sessions, where repeated exposure to items measuring the same, unidimensional construct resulted in an increase in correlations among items scores (Millar & Tessar, 1986). In fact, the effect increases with the number of items in a test (Knowles et al., 1996). Also, item response times decreased as the items appeared later in the test (Bargh, 1982). Together, these

findings indicate that an item response is determined not only by the content of the current item but the accumulated exposure to previous similar items.

Interest in this feature of the response process was further strengthened in the late 1980s and on into the 1990s with research by Knowles and colleagues. Knowles (1988) and Knowles and Byers (1996) found that item responses became more stable and correlated with the overall latent trait as they appeared later, sequentially, in a multi-item survey meant to measure locus of control. Similar results were found for tests of dogmatism, social desirability, (Knowles & Byers, 1996) and anxiety (Knowles, 1988; Knowles, Coker, Scott, Cook, & Neville, 1996), though the effect was stronger for tests in which the underlying trait being measured was less obvious to the respondents (e.g., locus of control versus anxiety, respectively). Hamilton and Shuminsky (1990) also uncovered the order effect in locus of control data. Specifically, the relationship between serial position and item-total correlation increased in an approximately linear fashion as in the above studies.

Interpretations of these findings are heavily steeped in cognitive theory. For example, as a respondent encounters more and more items reflecting the same characteristic or trait, self-schemas are more readily available in working memory storage. The aforementioned research (Hamilton and Shuminsky, 1990; Knowles, 1988; Knowles & Byers, 1996; Knowles et al., 1996) focused on the cognitive process entailed in rating one's self on personality traits, but differed from one another in terms of the particular stage of the response process affected. Knowles and colleagues emphasized the role of stage one (from Tourangeau & Rasinski's, 1998, framework): interpreting and understanding the content of the questionnaire, or meaning clarification (Knowles, 1988; Knowles et al., 1992; Knowles et al., 1996). In contrast, Hamilton and Shuminsky (1990) and others (e.g., Steinberg, 1994) were concerned with the second stage of the process: self-schema activation and the increase in threshold of repeated activation of the relevant self-schema.

With respect to the hypothesis of content knowledge, rather than self-awareness, driving the stability of scores, support was based on two studies. In the first study, there was a high serial order effect among groups who rated three different targets on a variety of traits: themselves, a best friend, or Bill Cosby (Knowles & Byers, 1996). The point of the experiment was to test whether trait ratings differed substantially across targets, and they did not. Had they differed substantially such that a serial effect was only occurring for self-ratings, then the results would support self-awareness. However, unless the respondent is privy to personal knowledge about, in particular, Bill Cosby, there should be no serial effect for ratings of him on a trait unless the respondent is becoming more familiar with the trait items and will become more consistent in how they choose to rate the person. In other words, the serial order effect was similar for one's self as well as an unknown other (e.g., Bill Cosby), suggesting that as the respondents encounter more items measuring the same trait, they become more familiar with what trait is being measured.

In a second study, respondents rated the degree to which a series of relevant versus random items reflect Locus of Control (versus other traits). The ratings became increasingly stable for items that appeared later in the test as if the respondents became more and more familiar with the trait content and which items measure which trait (Knowles & Byers, 1996). If these results were based on an experiment in which respondents rate themselves, rather than trying to define the trait, itself, the results could be explained by self-awareness. These latter two studies support the theory that item and test content familiarity, alone – rather than self-awareness - is driving the response.

Studies by both Hamilton and Shuminsky (1990) and Steinberg (1994) demonstrated empirical support for the hypothesis that level of self-awareness is largely responsible for the serial-order effect. Specifically, respondents were assigned to a high versus low self-awareness group. Self-awareness was manipulated by having respondents in the former group write a story about

themselves prior to completing the self-report (high self-awareness), whereas in the latter group, respondents wrote a story about another person such as George Washington (low self-awareness). The serial order effect was indeed stronger for respondents in the decreased self-focus group compared to the increased self-focus group. This phenomenon occurs because a respondent initially starts the test with little self-focus in comparison with respondents who are already focused, resulting in greater change (increased homogeneity) as the respondent becomes more and more self-reflective. Respondents who are already highly self-aware were more consistent in items responses from the beginning of the test session. Thus, these findings support the theory that, rather than familiarity with test content, it is self-concept, schemas, and perceptions of one's self drives the second stage of the response process; namely, comparison of self with the description given in the item stem.

From both points of view, attentional focus increasingly centers on the personality construct being measured, thereby influencing item response process. In other words either, or both, increased self-awareness and awareness of content contribute to a change in the response process. Specifically, the response process speeds up and becomes more streamlined. Additionally, estimates of the respondent's latent trait level become increasingly stable.

CHAPTER III

TECHNIQUES FOR MODELING CONTEXT EFFECTS: CLASSICAL VERSUS MODERN TEST THEORY

Classical Test Theory

The majority of studies investigating context effects in self-report personality responses have based their design around an analysis of reliability. The notion of reliability is founded in classical test theory (CTT), where observed test scores are a function of the true score and error (Lord & Novick, 1968):

$$Y_{pv} = \theta_p + \varepsilon_{pv} \quad (1)$$

where Y_{pv} is the observed test score of person *p* at sampling event *v*; θ_p is person *p*'s "true score" or average score across all *v* sampling events; and ε_{pv} refers to the random error associated with the sampling of person *p*'s behavior at the v^{th} sampling event. ε_{pv} is assumed to be unrelated to θ_p and is assumed to have a mean equal to zero in the population of sampling events for person *p*. This discrepancy between true and observed scores epitomizes the concept of reliability and has been addressed by indices for gauging it as well as the need to continuously replicate study results across multiple samples (Crocker & Algina, 1986).

Modern Test Theory

In the latter half of the 20^{th} century, a new measurement system emerged. This system has been referred to as modern test theory or *item* response theory (IRT) (Hambleton & Swaminathan, 1985; Lord & Novick, 1968). In IRT, precision of latent trait measurement differs from that of CTT in a number of ways. First, measurement characteristics are studied at the *item* level. This technique of focusing on the psychometric characteristics of individual test items logically corresponds to a greater capacity for improving the quality of the total test score

(Embretson & Reise, 2000). Second, traditional measures of reliability are not of primary concern because estimates are conditional upon a limited set of items. Indeed, there can be error in the estimation of a person's trait level, θ_p, but the focus in IRT modeling is on the amount of trait information that can be obtained from an item score given the associated item characteristics.

Finally, IRT differs from CTT in that model parameter interpretations are invariant to samples in the former but not the latter. Specifically, item parameters interpretations are invariant across respondents (with varying trait levels), and person parameter (trait level) interpretations are invariant across items used to measure the trait level of the respondent (Hambleton & Swaminathan, 1982; Lord & Novick, 1968). These invariant properties of parameter interpretations obtained from an IRT model make it a particularly attractive choice in investigating item-level phenomena in testing.

Use of CTT to Study the Order Effect

In psychometric studies of context effects, measurement of latent traits is typically centered on the total test score rather than individual items. In other words, it is this total test score, summed or averaged across items, that represents an individual's observed score (which is usually considered as a type of trait estimate). Item values take on meaning based on how they contribute to the total test score. Thus, in studies of context effects, the basis of analysis lies in the extent to which each individual item is correlated with the total test score. The order of items is manipulated, experimentally, via counterbalancing and using a Latin Square design. For example, in Knowles (1988), 30 items measuring Locus of Control appeared in each of 30 possible positions, resulting in 30 different test forms. The relative ordering of surrounding items was also counterbalanced such that a given item followed and preceded each of the other individual items approximately half the time.

The impact of order on reliability was tested as follows. First, the raw scores for each item were converted to z-scores across all respondents, regardless of order. Each person's observed (total) test score was based on an average of their 30 item z-scores. In order to target order effects, an additional set of z-scores were formed for each person by averaging the z-scores for all items in each position, regardless of the item content. As a result, each person had 30"position" z-scores and 30 test scores based on the average of z-scores for each item, regardless of position. Finally, the correlation between position z-scores and test scores was computed and found to be linearly related. For some scales, correlations ranged as much as .39 to .54 for positions one and 30, respectively. For other scales, however, smaller increases such as from .28 to .32 were revealed. This series of analytic steps was repeated in other studies of context effects, with similar results (Hamilton & Shuminsky, 1990; Knowles & Byers, 1996; Knowles et al., 1996).

The above CTT-oriented analysis is disadvantaged in the usual ways, relative to IRT. Namely, test scores are item-specific and inseparable from trait level reflected in the item (i.e., item location in IRT). Most importantly, however, is the focus on total score. The total score is inferred as the person's latent trait level, and the correlation between each position z-score and the total score (averaged z-score across items) is interpreted as consistency of items at the given position within the test. In other words, the item's identity is not distinguished from that of others. Rather, it is the position's identity that is of primary concern. In the studies previously mentioned, the content effect is unwanted and avoided by making sure that all items appear equally at each position. However, it would be advantageous to maintain the identity of and obtain trait information from individual items in order to distinguish content from context effects. An IRT analysis that includes sample-invariant parameters for both person and item would be ideal for this clarification. In addition, the impact of order on the precision of latent trait measurement provided by a given item (i.e., item discrimination) can also be determined.

Use of IRT to Study the Order Effect

There have been very few attempts to detect the impact of order effects on the precision of latent trait measurement in IRT. The one method used, to date, is referred to as differential item functioning (DIF) in which one or more item parameters vary across observed respondent groups. Local independence is assumed to be violated by a certain characteristic of the test taker such as gender, ethnicity, or situational context. For example, the item location of self-report personality items may be lower for a group that is instructed to "fake good" in comparison with a group that is instructed to "answer honestly" (Ferrando & Anguiano-Carrasco, 2009; Henry & Raju, 2006). Another scenario in which item location differs as a function of motivation involves job incumbent versus applicant pools (Robie, Zickar, & Schmit, 2001). The items for which psychometric characteristics systematically differ across situations demonstrate DIF. To date, only one study has used DIF to test the impact of order on the precision of item scores in measuring a unidimensional latent trait. Steinberg (1994) studied DIF among 20 Trait Anxiety items given in two, fixed order conditions. The responses were obtained with a Likert-type scale, so Samejima's graded response model (GRM) was fit to the data. The GRM is an IRT model for graded responses and is similar to the 2PL in that a discrimination parameter is included (Embretson & Reise, 2000; Samejima, 1969). Steinberg compared item discrimination parameters across conditions. DIF indeed occurred for discrimination on several items as a function of order condition, but no such variation occurred for item location. The implications of this early IRT study strengthens support for an order effect on the extent to which an item represents a trait total score, or a person's level of the latent trait. Moreover, the results are consistent with CTT studies in which mean item scores do not differ as a function of order. Therefore, further investigation in an IRT context is deemed warranted, as is the potential and usefulness of an IRT model that describes and quantifies this effect.

CHAPTER IV

CURRENT MODEL AND APPLICATION

The model proposed in the current study is unidimensional and polytomous in nature, meaning that not only is unidimensionality in the item response is expected, but the response formats for the test items contain more than two response categories. There are several polytomous IRT models to choose from, including Andrich's (1978) Rating Scale Model (RSM), Master's (1982) Partial Credit Model (PCM), Bock's (1972) Nominal Response Model (NRM), and Samejima's (1969) Graded Response Model (GRM). The model chosen for the current investigation is an extension of the Generalized Partial Credit Model (GPCM; Muraki, 1992), which is, itself, a generalization of Master's Partial Credit Model (PCM; 1982) to include a discrimination parameter that varies across items. This model was chosen based on support in both personality and educational literature suggesting better fit relative to the other polytomous models (Embretson & Reise, 2000; Zickar & Ury, 2002) in addition to research in which a competing model, the Graded Response Model (GRM; Samejima, 1969), has demonstrated poor fit (e.g., Chernyshenko, Stark, Chan, Drasgow, & Williams, 2001).

The Generalized Partial Credit Model

The GPCM is constructed to test the probability of person *p* selecting response category *x* from *c* possible response categories in item *i*.

$$P_{ix}(\theta_p) = \frac{exp\left[\sum_{j=0}^{x} a_i(\theta_p - \delta_{ij})\right]}{\sum_{r=0}^{c} exp\left[\sum_{j=0}^{r} a_i(\theta_p - \delta_{ij})\right]} \qquad 2$$

where $\sum_{j=0}^{0} a_i(\theta_p - \delta_{ij}) = 0$ and $\delta_{i0} = 0$ by definition.

Where a_i is the item discrimination (constant across all *j*=0, …, c categories for item *i*), and δ_{ij} is the category step parameter. The *j*th category step is the point on the latent trait continuum at

which the probability response function for category j+1 intersects with that of category j. Thus, δ_{ij} represents the point on the latent trait continuous at which the likelihood of person p selecting category j+1 equals that of selecting category j. For example, if a 5-point Likert response format is used in personality or attitude testing, there are 5 response categories, and 4 category steps,δ_{ij}. So, δ_{i3} represents the point on the latent trait continuum at which the probability of selecting "agree" (category 3) is equal relative to that of the preceding, second category (i.e., "neutral" or "I don't know"), and δ_{i4} corresponds to the point at which the selection of "strongly agree" (category 4) is equal to its preceding category "agree" (category 3). These step parameters are, therefore, coded zero through four. They are also easily identified when examining operating characteristic curves, as seen in Figure 1. These probability curves display the steps, or points along the latent trait continuum, at which a particular response category is more likely to be chosen. These steps, like the categorical responses to which they correspond, incrementally increase along the latent trait continuum. In other words, each step represents the point at which the likelihood of choosing one category equals that of choosing the previous category. Moreover, the higher one's level on the latent trait, the more likely one is to choose a higher category.

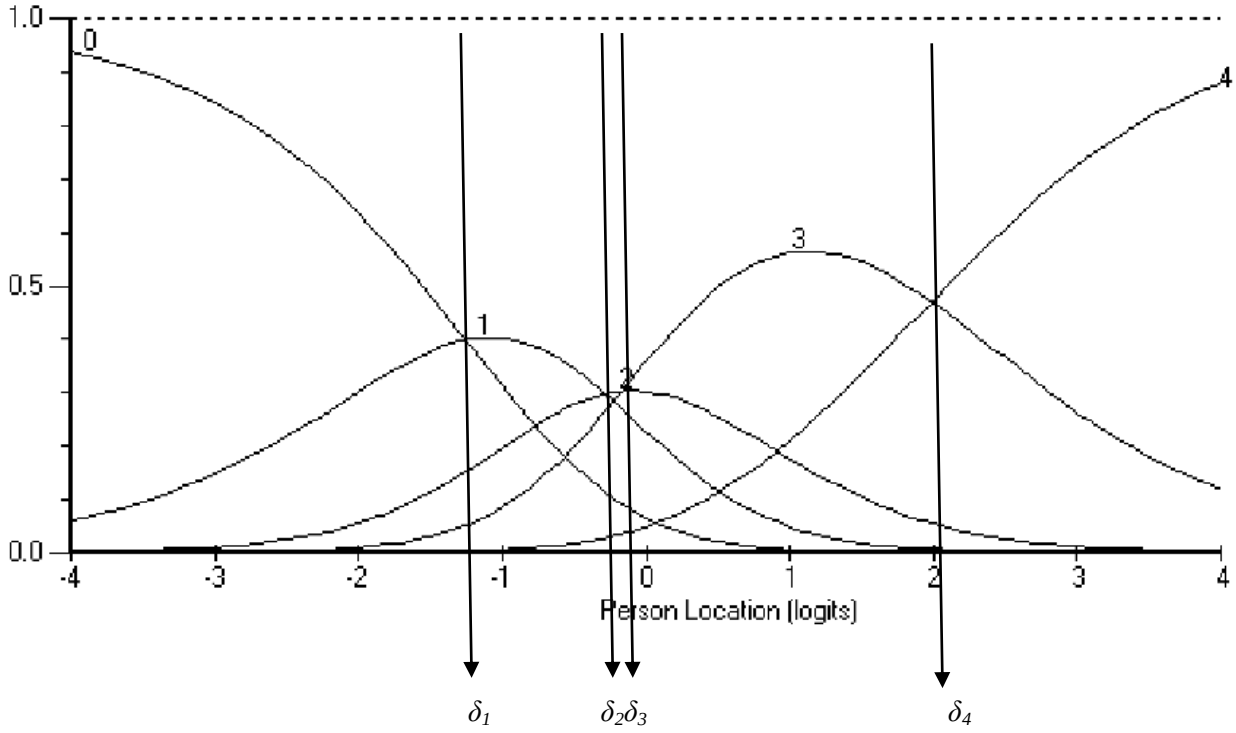


Figure 1: Operating Characteristic Curves for the Generalized Partial Credit Model

An alternative parameterization of the GPCM, as reflected in equation 2, can be seen in the following equation:

$$P_{ix}(\theta_p) = \frac{exp\left[\sum_{j=0}^{x} a_i(\theta_p - b_i + d_{ij})\right]}{\sum_{r=0}^{c} exp\left[\sum_{j=0}^{r} a_i(\theta_p - b_i + d_{ij})\right]} \text{ where } d_{1j} = 0 \qquad 3$$

Where δ_{ij} is now replaced with b_i - d_{ij}. b_i represents the *i*th's item location on the latent trait continuum, and d_{ji} is the category threshold for response category *j* in item *i*. The b_i and d_{ij} parameters can be converted to δ_{ij}, and vice-versa. For example, d_{ij} is the deviation of δ_{ij} from the item's overall location (b_i), while the average of δ_{ij} within an item is b_i (Muraki, 1992; Muraki, 1990). If the d_{ij}, as well as a_i, are equal across items (i.e., d_{ij} is replaced with d_j), the model becomes Andrich's (1978) Rating Scale model.

In educational and achievement testing, the GPCM indicates that a respondent must complete *c* steps in order to get full credit for an item. In other words, a researcher can use incrementally more correct answers as alternative response options, specifically for the purpose of ascertaining which stage in the task completion process an individual has reached. For example, in an algebra word problem requiring multiple stages in the solution process, the respondent may successfully execute some but not all of these stages. The degree of completion is, by definition in the model, interpreted as relevant to the respondent's latent trait level. In attitude and personality testing, however, each step "completed" reflects a higher level of the trait for the respondent. Thus, for a respondent, the more steps completed – i.e., the higher the response option selected – the greater the level of that respondent's latent trait.

The FACETS Model

The facets model (Linacre, 1989), or many-faceted Rasch model (MFRM; Linacre, 1993), can be used to take into account a variety of facets that impact an item response, the first two facets being person ability (theta) and item location (or difficulty). The facets model is originally in the family of Rasch models (Linacre, 1989), wherein items have a common slope (discrimination, e.g., a_i=1.0). Wang and Liu (2007) were the first to generalize the facets model to include both dichotomous and polytomous scoring as well as an item discrimination parameter that was allowed to vary across items. In polytomous models, the second facet can be a location of the step from category *j* to *j*+1.

Based on this model, the probability of choosing a particular response option, given a respondent's θ, is as follows:

$$P_{ix}(\theta_p) = \frac{exp\left[\sum_{j=0}^{x} \alpha_i(\theta_p - \delta_{ij}) - F_k\right]}{\sum_{r=0}^{c} exp\left[\sum_{j=0}^{r} \alpha_i(\theta_p - \delta_{ij}) - F_k\right]} \quad (4)$$

Where $\sum_{j=0}^{0} \alpha_i(\theta_p - \delta_{ij}) - F_k = 0$; $F_1 = 0$; and $\delta_{i0} = 0$ by definition

Same as with the GPCM, a_i is the item discrimination (constant across all j=0, …, c categories for item i), and δ_{ij} is the category step parameter. The jth category step is the point on the latent trait continuum at which the probability response function for category j+1 intersects with that of category j. An additional parameter is F_k , which represents the parameter value for the additional, third facet.

An alternative parameterization of the FACETS model that incorporates the GPCM, as reflected in equation 4, can be seen in the following equation:

$$P_{ix}(\theta_p) = \frac{exp\left[\sum_{j=0}^{x} a_i(\theta_p - b_i + d_{ij}) - F_k\right]}{\sum_{r=0}^{c} exp\left[\sum_{j=0}^{r} a_i(\theta_p - b_i + d_{ij}) - F_k\right]} \quad 5$$

where $d_{1j} = 0$ and F_1=0 by definition

Again, δ_{ij} is replaced with b_i - d_{ij}. b_i represents the ith's item location on the latent trait continuum, and d_{ji} is the category threshold for response category j in item i. Thus the b_i and d_{ij} parameters can be converted to δ_{ij}, and vice-versa (Muraki, 1992). Additionally, the parameter F_k is included to take into account an additional facet.

In most studies using the facets model, the third facet is rater severity where F is the degree of severity and k represents the rater who's scored the given item. Consequently, the facets model is often applied to constructed response test data, such as reading comprehension, which is scored by multiple raters (e.g., Linacre, 1999; Lunz, Wright, & Linacre, 1990). However, alternative facets can be applied such as a specific criterion on which the judges are rating item performance (Wang & Liu, 2007), or, as in the current study, item position.

The Generalized Partial Credit FACETS Model

The Generalized Partial Credit FACETS model adds parameters to Wang and Lui's model as follows. A parameter is added to reflect the 3rd facet, item position, albeit where position number

impacts discrimination. This GPCFM model, where the probability of a person with θ selects category *x* for item *i* ,is as follows :

$$P_{ix}(\theta_p) = \frac{exp\left[\sum_{j=0}^{x} E_k * \alpha_i(\theta_p - \delta_{ij}) - F_k\right]}{\sum_{r=0}^{c} exp\left[\sum_{j=0}^{r} E_k * \alpha_i(\theta_p - \delta_{ij}) - F_k\right]} \quad 6$$

Where $\sum_{j=0}^{0} \alpha_i(\theta_p - \delta_{ij}) - F_k = 0$; $F_1 = 0$; $E_1 = 1$; and $\delta_{i0} = 0$ by definition

where, a_i is the item discrimination and δ_{ij} is the category step parameter. The *j*th category step is the point on the latent trait continuum at which the probability response function for category *j*+1 intersects with that of category *j*. F_k represents the parameter value for a 3rd facet; namely, the influence of order on step parameters, and E_k reflects the 4th facet, which is the impact of order on discrimination.

An alternative parameterization of the GPCFM model, as reflected in equation 6, can be seen in the following equation:

$$P_{ix}(\theta_p) = \frac{exp\left[\sum_{j=0}^{x} E_k * a_i(\theta_p - b_i + d_{ij}) - F_k\right]}{\sum_{r=0}^{c} exp\left[\sum_{j=0}^{r} E_k * a_i(\theta_p - b_i + d_{ij}) - F_k\right]} \quad 7$$

where $d_{1j} = 0$, $F_1=0$, and $E_1=1$ by definition

Again, δ_{ij} is replaced with b_i - d_{ij}, where b_i represents the *i*th's item location on the latent trait continuum, and d_{ji} is the category threshold for response category *j* in item *i*. Additionally, the parameters F_k and E_k are included to take into account the order effect on steps and discrimination, respectively.

The purpose of this study was to develop a more sophisticated method of testing for and quantifying order effects in self-report personality data. Previous research has emphasized CTT and limited use of IRT methodology. Nevertheless, both methods have supported the presence of order effects in measurement scores for a variety of personality constructs. However, there has

been no IRT model constructed to more precisely target order effects, nor has the big five set of personality traits been investigated for this effect. Given the widespread use and importance of the big five, it would be ideal to apply a new model to this type of data.

The Generalized Partial Credit Facets Model (GPCFM) was constructed in the current project and applied to the investigation of context effects in self-report personality data when a polytomous response format was used. In the current study, it was expected that location or position of the item within the test would demonstrate a positive impact on item discrimination based on the collection of research previously described (Hamilton & Schuminsky, 1990; Knowles, 1988; Knowles & Byers, 1996; Ostrom et al., 1992; Steinberg, 1994). However, an impact of order on item location or steps was not expected based on the failure of the aforementioned research to find any effect of item position on mean item, or test, score.

In order to test the efficacy of the current model, two studies were conducted. In the first study, a series of simulations were performed in order to determine the conditions under which context effects are more accurately detected by the model. Information about the data demands associated with parameter recovery was also be obtained. In the second study, the model was applied to real data for the specific purpose of detecting context effects. This analysis was meant to illustrate how the GPCFM can be utilized in practice. Moreover, the efficacy of this model was further tested by comparing its fit to the data with nested models such as Muraki's (1992) GPCM and Wang and Liu's (2007) generalized FACETS model.

CHAPTER 5

METHOD, ANALYSIS, AND RESULTS

Simulation Design

All data was simulated using the SAS programming language. The design of the simulation study was as follows. The items were grouped into blocks based on order. For example, if there are 30 items in a test, the first ten items represented block 1, the second 10 items represented block 2, and the final 10 items represented the third and final block. However, item order was randomized for each participant. Thus, the items contained in each block were random as well.

Traditional factors were varied (e.g., sample size, test length), along with order effect parameters (per block) in the GPCFM. First, respondent sample size were small, medium, or large (500, 1,000, and 2000, respectively). Second, test length consisted of either 15 or 30 items. Third, the magnitude of e_k (order effect on item discrimination) was either small (e_2 =1.05, e_3=1.1) or large (e_2 =1.7, e_3=1.8). Fourth and finally, the magnitude of f_k (the order effect on item threshold) was either small (f_2 = .05 and f_3 =.1) or large (f_2 = .15 and f_3 =.2). Justification for the *e* and *f* parameter values will be discussed in the next section.

All 24 permutations of the latter four factors (3 x 2 x 2 x 2) were simulated and subsequently compared to estimated parameter values. Thirty independent data sets were drawn, randomly, resulting in 720 total simulated data samples (see Appendix A for SAS simulation code).

True Model Parameters and Data Generation

Determination of e_k magnitudes were based on an equation for converting polyserial item-total correlations into discrimination values (Van der Linden & Hambleton, 1997) and the range of item-total correlations uncovered in previous studies on order effects (Hamilton & Schuminsky,

1990; Knowles, 1988; Knowles & Byers, 1996; Ostrom et al., 1992). The equation for conversion is as follows:

$$\alpha_i = \frac{\rho_i}{\sqrt{1-\rho_i^2}} \quad (8)$$

Where α_i refers to the discrimination level for a given item, i, and ρ_i reflects the item-total polyserial correlation for a given item, i.

This formula was first used to convert the lowest and highest item-total correlations uncovered in the CTT-based studies of order effects to discrimination values. However, an important caveat must be stated. Because these item-total correlations are not polyserial, but rather pearson correlations, estimates were approximate and, if anything, lower than what would be expected if polyserial correlations were used. Thus, the discrimination values were approximated.

With respect to the aforementioned CTT studies, the study with the smallest observed order effect showed a range of correlations from .25 to .33 (Knowles, 1988), and the study with the largest effect ranged from .35 to .54 (Knowles, 1988). Next, the ratio of discrimination values converted from these correlation values was determined – namely, by dividing the higher discrimination value uncovered for items at the end of the test by the correlation of items appearing at the beginning of the test, as follows:

$$small\ e_k = \frac{.35}{.33} = 1.06$$

$$large\ e_k = \frac{.64}{.37} = 1.72$$

The ratio of the discriminations reflective of the smallest order effect served as a basis for selecting the set of small e_k parameters, whereas the ratio for the largest effects constituted the basis for the set of large e_k parameters.

The magnitude of f_k (the order effect on item step) were either small (f_2 = .05 and f_3 =.1) or large (f_2 = .15 and f_3 =.2). These values are based on previous serial order effect literature, in which mean differences were not found to be statistically significant (Hamilton & Schuminsky, 1990; Knowles, 1988; Knowles & Byers, 1996; Ostrom et al., 1992) nor was item location found to differ as a function of order (Steinberg, 1994).

In addition to generating values for the facet parameters, the remaining model parameters, θ_p, α_i, and δ_{ij} were determined based on the following methods. θ_p will be drawn from a random normal distribution, ~N(0, 1). The remaining item parameters were based on the results of an analysis of real data to which the GPCM has been applied, a technique similar to that which has been used in previous simulation research on the PCM and GPCM (Masters, 1982; Muraki, 1992). Specifically, the α_i, and δ_{ij} parameters were sampled from those uncovered in an analysis of responses to a 33-item, four-category response anxiety inventory with a sample size that exceeded 2,000 (Walter et al., 2007). The α_i values ranged from .83 to 2.6, the δ_{i1} values ranged from -2.81 to -.02, the δ_{i2} values ranged from -1.56 to 1.71, and finally the δ_{i3} parameters ranged from -.18 to 3.30. Sets of parameters that are associated with one of the 33 items were randomly assigned, with replacement, to each item in the simulation data set.

Estimation of the Model Parameters

Model Identification

In order to ensure that the GFPCM was identified during estimation, three steps were taken. First, the measurement scale was fixed for θ as follows: $\theta \sim N(0,1)$, as is common in standard IRT practices. Second, consistent with the GPCM, the first category in each block was fixed to be 0, as follows: $\sum_{j=0}^{0} \alpha_i(\theta_p - \delta_{ij}) = 0$. Thus, the weighted, α_i, distance between θ_p and δ_{i2} as well as

the distance between θ_p and δ_{i3} and so on for all remaining thresholds, was allowed to vary across items. Finally, the new parameters, which must be allowed to vary across blocks, were fixed as follows: $F = 0$ for block one but were allowed to vary across blocks two and three. Also, by fixing *F* to 0 in the first block, we are able to see how the order of items (represented by block) results in an increase or decrease in the item threshold, or conversely, the item step. Similarly, *E* was fixed to 1 in the first block in order to allow α_i to remain unaffected and see how it may change as blocks (order) increase. Thus, *E* was allowed to vary across blocks two and three, only.

Markov Chain Monte Carlo

In all analyses, the parameters of the new model were estimated via the WinBUGS (Bayesian inference Using Gibbs Sampling) computer program in which a Markov Chain Monte Carlo (MCMC) estimation algorithm is implemented (Spiegelhalter, Thomas, Best, & Lunn, 2003). Monte Carlo refers to a stochastic process or successive steps in a "random walk" within a Markov Chain. The Markov Chain, itself, refers to a sequence of random variables that are ultimately sampled from a posterior distribution. Indeed, the ultimate goal in MCMC is for the chain to reach a stationary distribution, which is equivalent to the posterior distribution (Kim & Bolt, 2007; Patz & Junker, 1999a; Patz & Junker, 1999b). The number of states, or iterations, in this chain that are required to reach a stationary distribution is somewhat variable, depending on the algorithm or sampling method chosen and other factors such as number of items and people. However, in general, if there are *t* states in the chain, it is likely that the stationary distribution will become increasingly likely and thus approach the posterior distribution as *t* increases.

Once the stationary and posterior distributions become approximately equal, simulated observations (states) can be sampled from the chain to make inferences about model parameters (Gelman & Rubin, 1992; Kim & Bolt, 2007; Patz & Junker, 1999a; Patz & Junker, 1999b). The

following equations represent posterior distributions for each of the parameters in the GPCFM model:

$$p(\theta|X,\alpha,\delta_1,\delta_2,\delta_3,e_2,e_3,f_2,f_3) \propto p(X|\theta,\alpha,\delta_1,\delta_2,\delta_3,e_2,e_3,f_2,f_3)p(\theta) \quad 9$$

$$p(\alpha|X,\theta,\delta_1,\delta_2,\delta_3,e_2,e_3,f_2,f_3) \propto p(X|\theta,\alpha,\delta_1,\delta_2,\delta_3,e_2,e_3,f_2,f_3)p(\alpha) \quad 10$$

$$p(\delta_1|X,\theta,\alpha,\delta_2,\delta_3,e_2,e_3,f_2,f_3) \propto p(X|\theta,\alpha,\delta_1,\delta_2,\delta_3,e_2,e_3,f_2,f_3)p(\delta_1) \quad 11$$

$$p(\delta_2|X,\theta,\alpha,\delta_1,\delta_3,e_2,e_3,f_2,f_3) \propto p(X|\theta,\alpha,\delta_1,\delta_2,\delta_3,e_2,e_3,f_2,f_3)p(\delta_2) \quad 12$$

$$p(\delta_3|X,\theta,\alpha,\delta_1,\delta_2,e_2,e_3,f_2,f_3) \propto p(X|\theta,\alpha,\delta_1,\delta_2,\delta_3,e_2,e_3,f_2,f_3)p(\delta_3) \quad 13$$

$$p(e_2|X,\theta,\alpha,\delta_1,\delta_2,\delta_3,e_3,f_2,f_3) \propto p(X|\theta,\alpha,\delta_1,\delta_2,\delta_3,e_2,e_3,f_2,f_3)p(e_2) \quad 14$$

$$p(e_3|X,\theta,\alpha,\delta_1,\delta_2,\delta_3,e_2,f_2,f_3) \propto p(X|\theta,\alpha,\delta_1,\delta_2,\delta_3,e_2,e_3,f_2,f_3)p(e_3) \quad 15$$

$$p(f_2|X,\theta,\alpha,\delta_1,\delta_2,\delta_3,e_2,e_3,f_3) \propto p(X|\theta,\alpha,\delta_1,\delta_2,\delta_3,e_2,e_3,f_2,f_3)p(f_2) \quad 16$$

$$p(f_3|X,\theta,\alpha,\delta_1,\delta_2,\delta_3,e_2,e_3,f_2) \propto p(X|\theta,\alpha,\delta_1,\delta_2,\delta_3,e_2,e_3,f_2,f_3)p(f_3) \quad 17$$

In the above formulas, the likelihoods and prior probabilities that fall on the right side of the equation are, collectively, the numerator in a more complex equation. The denominators for the θ and α posterior distributions, above, are as follows:

$$\int p(X|\theta,\alpha,\delta_1,\delta_2,\delta_3,e_2,e_3,f_2,f_3)p(\theta)p(\alpha)p(\delta_1),p(\delta_2),p(\delta_3),p(e_2),p(e_3),p(f_2),p(f_3)d\theta \quad 18$$

$$\int p(X|\theta,\alpha,\delta_1,\delta_2,\delta_3,e_2,e_3,f_2,f_3)p(\theta)p(\alpha)p(\delta_1),p(\delta_2),p(\delta_3),p(e_2),p(e_3),p(f_2),p(f_3)d\alpha \quad 19$$

When considering remaining parameters, this integral differs from the above in that the parameter of interest is listed at the end of the equation, after *d*. The above components of the posterior distribution are essentially normalizing constants that result in closed-form, full conditional distributions. These full conditionals, which are mathematically arduous and not always feasible

for sampling, complicate the process of drawing samples from the posterior distribution. Therefore, additional algorithms have been developed for facilitating and simplifying the sampling process from state to state along the Markov chain, and these algorithms have been successfully applied to a variety of IRT models (Patz & Junker, 1999a; 1999b).

The Gibbs sampler is perhaps the most common algorithm applied in MCMC and solves some of the computation difficulties associated with the full conditional distributions. The latter is accomplished by iteratively conditioning upon known parameter values from the previous chain state in order to determine values for remaining parameters in a subsequent state in the chain. Thus, the process can be described as "divide-and-conquer," albeit within chains and involving sampled states, due to the emphasis on conditional probabilities for each draw. For example, suppose a researcher wants to iteratively determine all of the parameter values for each state in the chain. Only the most recent state value is applicable; previous samples do not impact current samples. Thus, each chain state value is determined by its conditional probability given the remaining parameters and their previous, *t*-1, state values. A set of draws for each model parameter will proceed as follows:

$$\theta_1^t \sim p(\theta_1 | \theta_2^{t-1}, \dots, \theta_P^{t-1}, X, \alpha_1^{t-1}, \dots, \alpha_I^{t-1}, \delta_{11}^{t-1}, \dots, \delta_{I1}^{t-1}, \delta_{12}^{t-1}, \dots, \delta_{I2}^{t-1}, \delta_{13}^{t-1}, \dots, \delta_{I3}^{t-1}, e_2^{t-1}, e_3^{t-1}, f_2^{t-1}, f_3^{t-1}) \quad (20)$$

...

$$\theta_P^t \sim p(\theta_P | \theta_1^t, \dots, \theta_{P-1}^t, X, \alpha_1^{t-1}, \dots, \alpha_I^{t-1}, \delta_{11}^{t-1}, \dots, \delta_{I1}^{t-1}, \delta_{12}^{t-1}, \dots, \delta_{I2}^{t-1}, \delta_{13}^{t-1}, \dots, \delta_{I3}^{t-1}, e_2^{t-1}, e_3^{t-1}, f_2^{t-1}, f_3^{t-1}) \quad (21)$$

$$\alpha_1^t \sim p(\alpha_1 | \alpha_2^{t-1}, \dots, \alpha_I^{t-1}, \theta_1^t, \dots, \theta_P^t, X, \delta_{11}^{t-1}, \dots, \delta_{I1}^{t-1}, \delta_{12}^{t-1}, \dots, \delta_{I2}^{t-1}, \delta_{13}^{t-1}, \dots, \delta_{I3}^{t-1}, e_2^{t-1}, e_3^{t-1}, f_2^{t-1}, f_3^{t-1}) \quad (22)$$

…

$$\alpha_I^t \sim p(\alpha_I | \alpha_1^t, \ldots, \alpha_{I-1}^t, \theta_1^t, \ldots, \theta_P^t, X, \delta_{11}^{t-1}, \ldots, \delta_{I1}^{t-1}, \delta_{12}^{t-1}, \ldots, \delta_{I2}^{t-1}, \delta_{13}^{t-1}, \ldots, \delta_{I3}^{t-1}, e_2^{t-1}, e_3^{t-1}, f_2^{t-1}, f_3^{t-1})$$

23

$$\delta_{11}^t \sim p(\delta_{11} | \delta_{12}^{t-1}, \ldots, \delta_{I1}^{t-1}, \theta_1^t, \ldots, \theta_P^t, \alpha_1^t, \ldots, \alpha_I^t, X, \delta_{12}^{t-1}, \ldots, \delta_{I2}^{t-1}, \delta_{13}^{t-1}, \ldots, \delta_{I3}^{t-1}, e_2^{t-1}, e_3^{t-1}, f_2^{t-1}, f_3^{t-1})$$

24

…

$$\delta_{I1}^t \sim p(\delta_{I1} | \delta_{11}^t, \ldots, \delta_{I1-11}^t, \theta_1^t, \ldots, \theta_P^t, \alpha_1^t, \ldots, \alpha_I^t, X, \delta_{12}^{t-1}, \ldots, \delta_{I2}^{t-1}, \delta_{13}^{t-1}, \ldots, \delta_{I3}^{t-1}, e_2^{t-1}, e_3^{t-1}, f_2^{t-1}, f_3^{t-1})$$

25

$$\delta_{12}^t \sim p(\delta_{12} | \delta_{12}^{t-1}, \ldots, \delta_{I2}^{t-1}, \theta_1^t, \ldots, \theta_P^t, \alpha_1^t, \ldots, \alpha_I^t, \delta_{11}^t, \ldots, \delta_{I1}^t, X, \delta_{13}^{t-1}, \ldots, \delta_{I3}^{t-1}, e_2^{t-1}, e_3^{t-1}, f_2^{t-1}, f_3^{t-1})$$

26

…

$$\delta_{I2}^t \sim p(\delta_{I2} | \delta_{12}^t, \ldots, \delta_{I2-12}^t, \theta_1^t, \ldots, \theta_P^t, \alpha_1^t, \ldots, \alpha_I^t, \delta_{11}^t, \ldots, \delta_{I1}^t, \delta_{13}^{t-1}, \ldots, \delta_{I3}^{t-1}, e_2^{t-1}, e_3^{t-1}, f_2^{t-1}, f_3^{t-1})$$

27

$$\delta_{13}^t \sim p(\delta_{13} | \delta_{13}^{t-1}, \ldots, \delta_{I3}^{t-1}, \theta_1^t, \ldots, \theta_P^t, \alpha_1^t, \ldots, \alpha_I^t, \delta_{11}^t, \ldots, \delta_{I1}^t, X, \delta_{12}^t, \ldots, \delta_{I2}^t, e_2^{t-1}, e_3^{t-1}, f_2^{t-1}, f_3^{t-1})$$

28

…

$$\delta_{I3}^t \sim p(\delta_{I3} | \delta_{13}^t, \ldots, \delta_{I3-13}^t, \theta_1^t, \ldots, \theta_P^t, \alpha_1^t, \ldots, \alpha_I^t, \delta_{11}^t, \ldots, \delta_{I1}^t, \delta_{12}^t, \ldots, \delta_{I2}^t, e_2^{t-1}, e_3^{t-1}, f_2^{t-1}, f_3^{t-1})$$

29

$$e_2^t \sim p(e_2 | e_2^{t-1}, \theta_1^t, \ldots, \theta_P^t, \alpha_1^t, \ldots, \alpha_I^t, \delta_{11}^t, \ldots, \delta_{I1}^t \delta_{12}^t, \ldots, \delta_{I2}^t, \delta_{13}^t, \ldots, \delta_{I3}^t, X, e_3^{t-1}, f_2^{t-1}, f_3^{t-1})$$

30

$$e_3^t \sim p(e_3 | e_3^{t-1}, \theta_1^t, \ldots, \theta_P^t, \alpha_1^t, \ldots, \alpha_I^t, \delta_{11}^t, \ldots, \delta_{I1}^t \delta_{12}^t, \ldots, \delta_{I2}^t, \delta_{13}^t, \ldots, \delta_{I3}^t, e_2^t, X, f_2^{t-1}, f_3^{t-1})$$

31

$$f_2^t \sim p(f_2 | f_2^{t-1}, \theta_1^t, \ldots, \theta_P^t, \alpha_1^t, \ldots, \alpha_I^t, \delta_{11}^t, \ldots, \delta_{I1}^t \delta_{12}^t, \ldots, \delta_{I2}^t, \delta_{13}^t, \ldots, \delta_{I3}^t, e_2^t, e_3^t, f_3^{t-1}) \quad (32)$$

$$f_3^t \sim p(f_3 | f_3^{t-1}, \theta_1^t, \ldots, \theta_P^t, \alpha_1^t, \ldots, \alpha_I^t, \delta_{11}^t, \ldots, \delta_{I1}^t \delta_{12}^t, \ldots, \delta_{I2}^t, \delta_{13}^t, \ldots, \delta_{I3}^t, e_2^t, e_2^t, f_2^t) \quad (33)$$

Although the above sampling process is ideal, it may also become cumbersome because not all *t* or *t*-1 values for all parameters, on which each model parameter is conditioned, may be known. Therefore, even more simplified algorithms have been developed in which proposal distributions or candidate steps are used to compute a probability of moving from one state to another in a chain, thereby determining what the value of a parameter will be in the next state of the chain. Examples include Metropolis-Hastings (Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953) and a hybrid approach called Metropolis-Hastings within Gibbs (Patz & Junker, 1999). Various algorithms are specifically utilized in WinBUGS. For example, in the current model, the following parameters were estimated based on the following sampling algorithms. Parameters $F, \delta_1, \delta_2, \delta_3,$ and θ were submitted to adaptive metropolis, an algorithm that is intended to "learn" how to better sample based on proposal distributions (Harrio, Saksman, & Tamminen, 2001), while E and α were submitted to slice sampling, in which rejection sampling is used but there is no need to manually tune to the candidate function (Neal, 1997).

Given the nature of the Bayesian, MCMC estimation process, prior distributions are specified for all estimated parameters. The following prior distributions were used:

$$\theta_p \sim N(0,1) \quad (34)$$

$$\alpha_i \sim logN(0,.25) \quad (35)$$

$$\delta_{ij} \sim N(0,4) \quad (36)$$

$$e_k \sim logN(0,.25) \quad (37)$$

$$f_k \sim N(0,4) \quad (38)$$

These particular priors were chosen based on priors typically used in Item Response Theory software such as Parscale (Muraki & Bock, 2002) and prior research in which these distributions were shown to be most useful during the estimation process (e.g., Kim & Bolt, 2007). Please see Appendix B for a sample of WinBUGS code used to estimate the GPCFM.

At least ten thousand iterations (i.e., total MCMC iterations, interchangeably referred to as samples) were performed for each of the 720 simulations. In order to ensure convergence of the model to the intended posterior distribution, pilot research was conducted (see output Figures 2a through 2e and description of procedure and analysis in Appendix C). Based on these results, the first 1000 iterations were discarded, thus referred to as "burn-in" iterations, thinning was used but differed as a function of sample size (for n=500, every 5th iteration was retained; for n=1000 and n=2000, every 3rd iteration was retained). In addition, the total number of iteration retained differed as a function of sample size. For n=500, 20,000 iterations were retained, whereas for n=1000 and n=2000, only 10,000 iterations were required and retained. The reason for differing the number of total iterations was based on the Geweke's criterion (1992) computed for the outcome of pilot research. Namely, the results indicated that more iterations were to reach convergence when n=500 relative to n=1000 and n=2000 – i.e., 20,000 for n=500 and 10,000 for n=1000 and n=2000. Thus, in order to attain convergence, these iterations constituted the sample data used to compute estimates of model parameters across the different conditions (e.g., sample size factor).

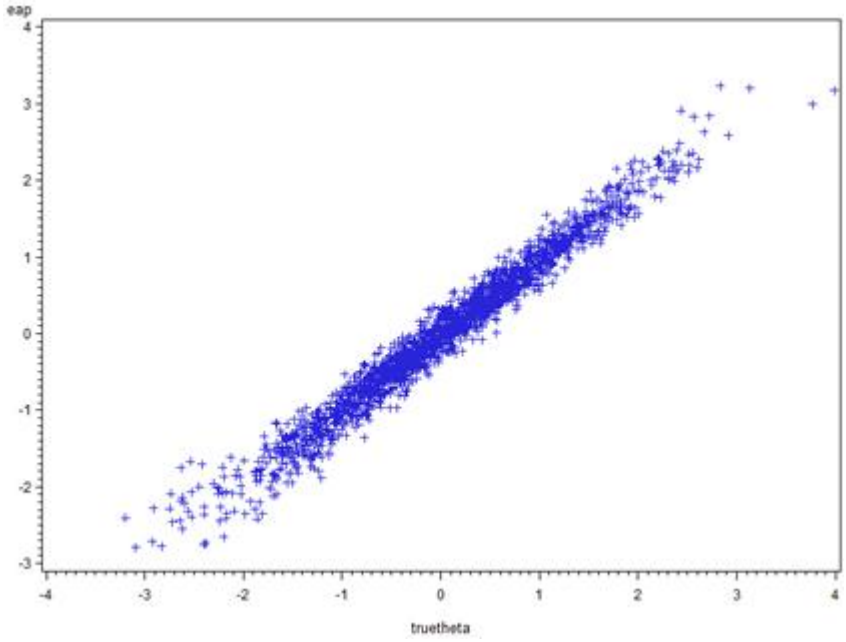


Figure 2a: True versus Estimated Latent Trait Values for Simulation Pilot Study Results

r=.984**

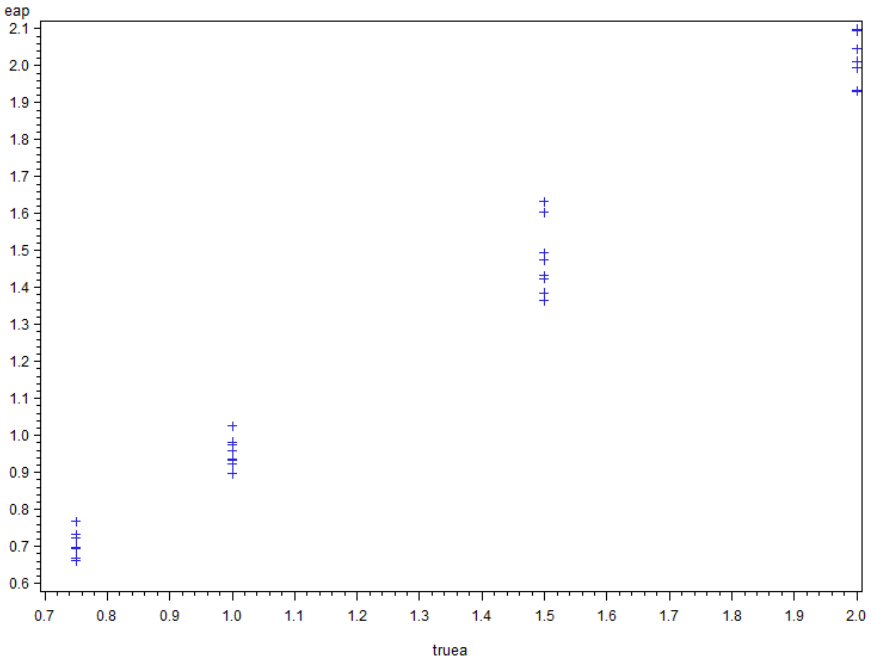


Figure 2b: True versus Estimated Item Discrimination Values for Simulation Pilot Study Results

r=.992**

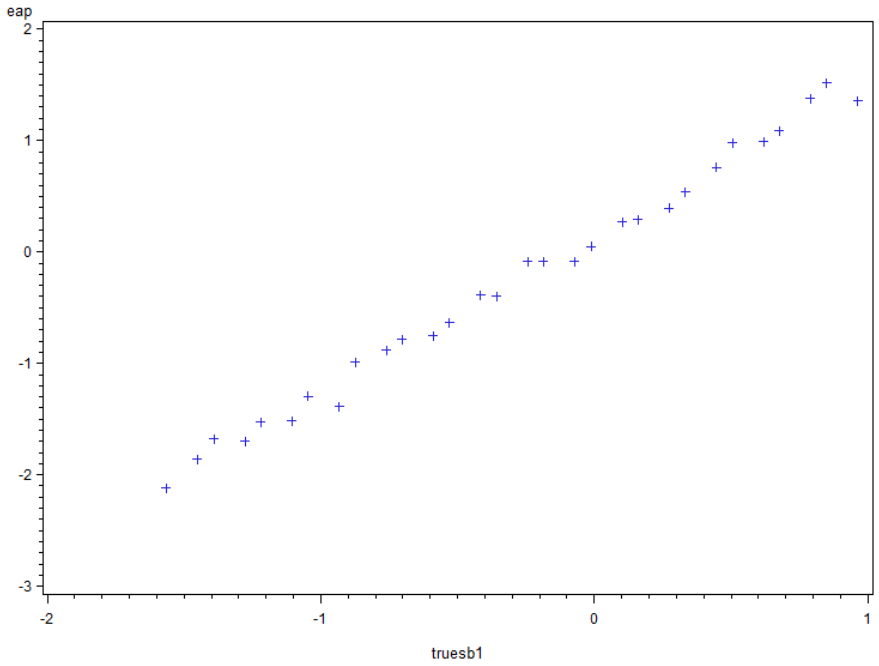


Figure 2c: True versus Estimated Step One Values for Simulation Pilot Study Results

r=.996**

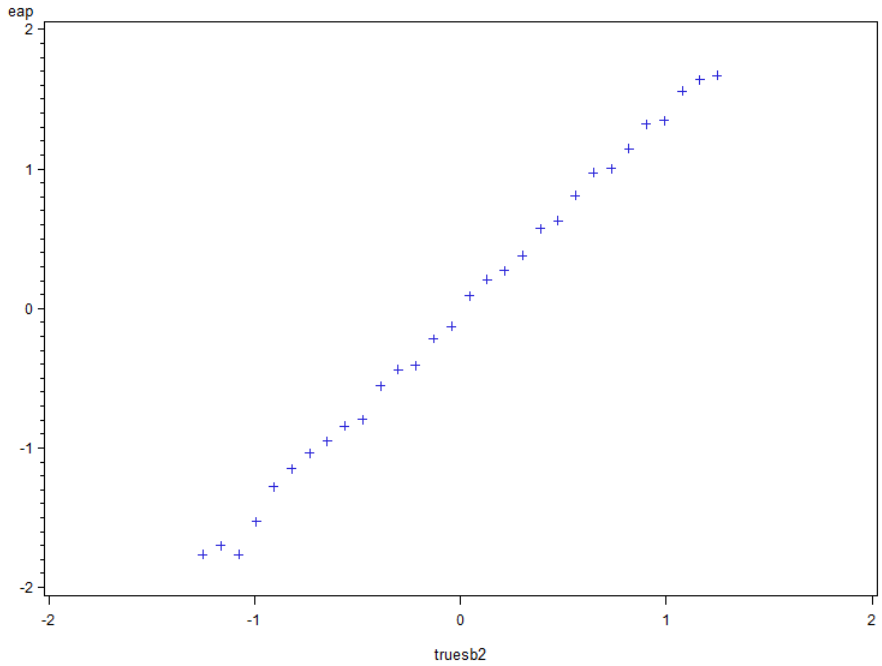


Figure 2d: True versus Estimated Step Two Values for Simulation Pilot Study Results

r=.999**

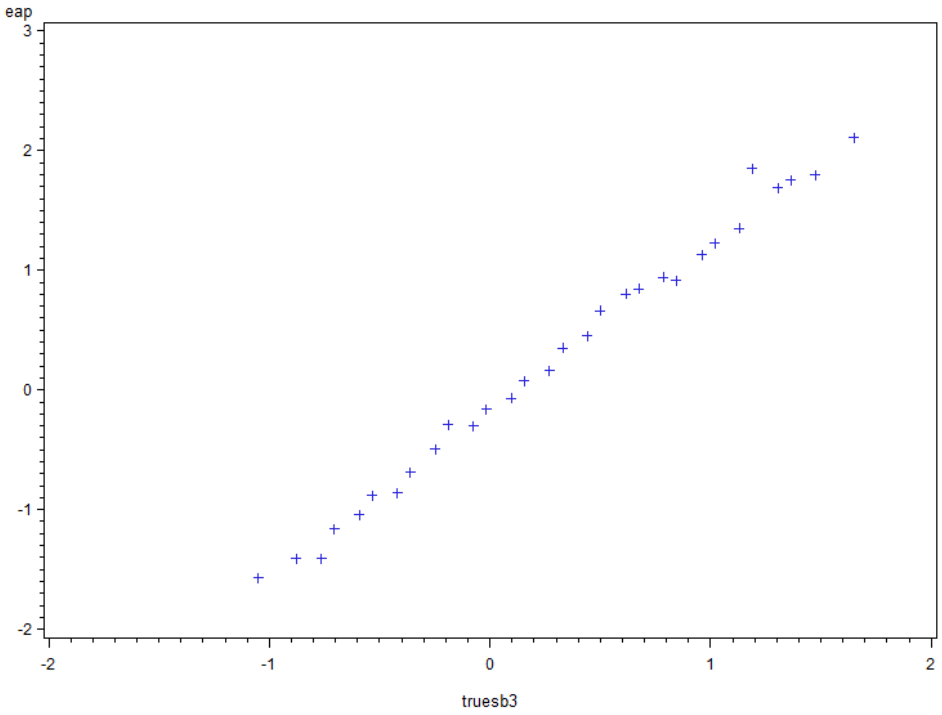


Figure 2e: True versus Estimated Step Three Values for Simulation Pilot Study Results

r=.997**

Convergence

Convergence of the MCMC estimation process was tested in two ways: 1) visual examination of trace plots, and 2) statistical tests. Trace plots were examined for homogeneity in mean values across iterations for each parameter within a given simulation and its associated condition. A set of trace plots for each parameter from a simulation performed in two conditions can be found in Appendix D. The two conditions represent the expected best and worst case scenario in terms of the effectiveness of parameter recovery – namely, n=2000, i=30, small order effects and n=500, i=15, and large order effects, respectively. It can be seen that the trace plots demonstrated convergence because the sample values vary little around the mean parameter value across iterations. This was the finding for all parameters, regardless of condition. Next, a series of convergence indices appropriate for MCMC methods were computed. These methods included

Geweke's criterion (1992) and Raftery and Lewis's criterion (1992), both tested in the BOA program (Spiegelhalter, Best, Carlin, & Van der Linden, 2002). Specifically, CODA information (chain values) was extracted from WinBUGS output and analyzed via the BOA (Bayesian Output Analysis) programming software. Geweke's (1992) criterion is composed of the difference between the mean of the first 10% of sampled values and the mean of the last 50% of sampled values, divided by pooled standard deviation. The output for this criterion is substantial and too large to include in this paper (i.e., a test is performed on each parameter in each condition), though the output values may be obtained from the author upon request. Nevertheless, a thorough examination of these values revealed that the majority (i.e., approximately 95%) of the parameters, regardless of condition, met the above requirement (non-significance of the difference between the first 10% and last 50% of sampled values); thus, convergence is inferred.

Raftery and Lewis's (1992) technique involved determining the number of samples required to effectively estimate the posterior (i.e., greater precision). The resulting output from BOA informs the researcher the total number of samples needed, along with the number of burn-ins required (i.e., the samples to be thrown out), the number of samples that need to be thinned (every t^{th} sample after the burn-in sequence should be retained due to potential issues arising from autocorrelations). Given that pilot simulations were run and tested with this set of criteria, it is not surprising that convergence was reached. Namely, the results of analyses were based on the criteria set forth during pilot testing (see page 30, paragraph 2). Moreover, these results served as a guide for determining the number of iterations and amount of thinning as previously described on page 30.

Analysis and Results for Simulation Study

In order to determine the extent to which model parameters were effectively recovered from the estimation process, a series of steps were performed. First, discrepancies in true (simulated) and mean parameter estimates across simulations within a condition, and across conditions, were examined. Moreover, the following error measurements were obtained and examined closely for each item parameter within each condition: standard deviation (SD), MCMC standard error (SE), and Root Mean Squared Error (RMSE). The SD for each parameter is based on the differences between replication estimates across the 30 replications. That is, the difference between the mean estimate for the condition and the individual estimates in the replication. The MCMC standard errors are computed using the posterior standard deviations (PSD), where the PSD for each parameter represents the square root of the average variance of the iterations (i.e., samples) after the burn-in. Finally, RMSE represents the sum of the square root of the squared differences between the true parameter value and the estimated parameter values across replications in a condition, divided by the number of samples (i.e., 30).

The SDs, MCMC SEs, and RMSEs, and mean parameter value estimates for each item in the simulation condition (i.e., across the 30 simulations) as well as averages across each of the 24 conditions can be found in Appendix E. Summary (i.e., averages across items) of these error measurements, including bias (i.e., true versus estimated parameter values, for each conditions can be found in Tables 1a through 4b, where each table refers to a particular error measurement type so that a) values for a particular parameter can be compared across conditions and b) values can be compared between parameters to determine if any parameters were recovered more or less effectively than other parameters.

Table 1a: Summary of SDs for MCMC Parameter Estimates across Items for i=15

Sample size	Order effect condition	α	δ_1	δ_2	δ_3	e	f
500	Small e, small f	.14	.12	.10	.11	.01	.01
	Small e, large f	.13	.13	.11	.11	.01	.01
	Large e, small f	.14	.13	.12	.11	.01	.01
	Large e, Large e	.12	.14	.09	.11	.01	.01
1000	Small e, small f	.07	.07	.07	.08	.01	.01
	Small e, large f	.09	.09	.07	.09	.01	.01
	Large e, small f	.09	.09	.08	.07	.004	.003
	Large e, Large e	.09	.08	.06	.08	.003	.003
2000	Small e, small f	.06	.05	.06	.06	.002	.002
	Small e, large f	.07	.06	.04	.06	.002	.001
	Large e, small f	.07	.06	.06	.04	.001	.001
	Large e, Large e	.07	.06	.05	.05	.001	.002

Note: SD = Variation of iteration values within a chain, across simulations

Table 1b: Summary of SDs for MCMC Parameter Estimates across Items for i=30

Sample size	Order effect condition	α	δ_1	δ_2	δ_3	e	F
500	Small e, small f	.11	.14	.11	.14	.01	.01
	Small e, large f	.11	.14	.10	.13	.01	.01
	Large e, small f	.11	.13	.09	.12	.01	.01
	Large e, Large e	.10	.14	.11	.11	.01	.01
1000	Small e, small f	.07	.10	.09	.08	.004	.01
	Small e, large f	.08	.10	.08	.09	.004	.01
	Large e, small f	.09	.08	.07	.08	.01	.004
	Large e, Large e	.08	.08	.09	.08	.004	.002
2000	Small e, small f	.04	.08	.07	.06	.002	.002
	Small e, large f	.04	.06	.05	.04	.001	.002
	Large e, small f	.05	.05	.05	.05	.001	.001
	Large e, Large e	.06	.0	.06	.06	.001	.001

Note: SD = Variation of iteration values within a chain, across simulations

Table 2a: Summary of SEs for MCMC Parameter Estimates across Items for i=15

Sample size	Order effect condition	α	δ_1	δ_2	δ_3	e	f
500	Small e, small f	.15	.14	.11	.12	.07	.08
	Small e, large f	.15	.14	.11	.12	.09	.09
	Large e, small f	.15	.13	.10	.10	.08	.08
	Large e, Large e	.15	.12	.09	.10	.07	.07
1000	Small e, small f	.11	.09	.07	.08	.07	.07
	Small e, large f	.11	.09	.08	.08	.07	.06
	Large e, small f	.11	.08	.07	.07	.07	.05
	Large e, Large e	.11	.08	.06	.07	.05	.05
2000	Small e, small f	.08	.06	.06	.05	.05	.06
	Small e, large f	.08	.06	.04	.06	.05	.05
	Large e, small f	.09	.05	.04	.04	.04	.03
	Large e, Large e	.08	.05	.04	.05	.04	.04

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Table 2b: Summary of SEs for MCMC Parameter Estimates across Items for i=30

Sample size	Order effect condition	α	δ_1	δ_2	δ_3	e	f
500	Small e, small f	.12	.10	.13	.13	.08	.07
	Small e, large f	.11	.15	.14	.14	.09	.07
	Large e, small f	.11	.13	.10	.11	.07	.09
	Large e, Large e	.11	.12	.11	.11	.07	.08
1000	Small e, small f	.09	.10	.09	.09	.07	.06
	Small e, large f	.08	.10	.09	.09	.07	.06
	Large e, small f	.08	.09	.08	.08	.05	.06
	Large e, Large e	.08	.08	.07	.07	.06	.05
2000	Small e, small f	.06	.07	.06	.06	.05	.06
	Small e, large f	.06	.07	.06	.06	.05	.05
	Large e, small f	.06	.06	.05	.06	.04	.04
	Large e, Large e	.06	.05	.05	.05	.05	.04

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Table 3a: Summary of RMSEs for MCMC Parameter Estimates across Items for i=15

Sample size	Order effect condition	α	δ_1	δ_2	δ_3	E	f
500	Small e, small f	.16	.16	.11	.11	.08	.07
	Small e, large f	.15	.16	.11	.11	.08	.09
	Large e, small f	.14	.19	.13	.11	.09	.08
	Large e, Large e	.13	.17	.10	.12	.08	.07
1000	Small e, small f	.07	.09	.08	.08	.07	.06
	Small e, large f	.09	.09	.08	.09	.07	.07
	Large e, small f	.09	.10	.08	.07	.06	.08
	Large e, Large e	.10	.09	.07	.08	.06	.07
2000	Small e, small f	.06	.06	.07	.06	.05	.04
	Small e, large f	.07	.06	.05	.06	.04	.04
	Large e, small f	.07	.07	.07	.04	.03	.05
	Large e, Large e	.09	.07	.06	.05	.05	.03

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table 3b: Summary of RMSEs for MCMC Parameter Estimates across Items for i=30

Sample size	Order effect condition	α	δ_1	δ_2	δ_3	e	f
500	Small e, small f	.17	.19	.12	.14	.09	.07
	Small e, large f	.18	.19	.11	.13	.08	.08
	Large e, small f	.17	.19	.10	.12	.07	.09
	Large e, Large e	.17	.19	.13	.11	.07	.07
1000	Small e, small f	.10	.12	.10	.08	.05	.05
	Small e, large f	.11	.12	.08	.09	.06	.05
	Large e, small f	.12	.11	.08	.08	.05	.07
	Large e, Large e	.11	.10	.09	.08	.07	.07
2000	Small e, small f	.05	.09	.07	.06	.05	.03
	Small e, large f	.06	.07	.05	.05	.03	.04
	Large e, small f	.07	.05	.06	.05	.03	.03
	Large e, Large e	.07	.06	.06	.06	.03	.04

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table 4a: Summary of Bias for MCMC Parameter Estimates across Items for i=15

Sample size	Order effect condition	α	δ_1	δ_2	δ_3	e	f
500	Small e, small f	.08	.12	.05	.02	.03	.02
	Small e, large f	.08	.10	.02	.02	.02	.02
	Large e, small f	.05	.14	.05	.04	.03	.03
	Large e, Large e	.06	.10	.05	.05	.02	.02
1000	Small e, small f	.001	.06	.04	.02	.02	.01
	Small e, large f	.001	.04	.04	.01	.01	.01
	Large e, small f	.01	.06	.04	.03	.02	.02
	Large e, Large e	.01	.05	.04	.02	.01	.01
2000	Small e, small f	.03	.04	.04	.001	.01	.004
	Small e, large f	.02	.03	.04	.001	.004	.003
	Large e, small f	.03	.04	.04	.02	.004	.003
	Large e, Large e	.03	.04	.04	.02	.004	.004

Note: Bias = Difference Between True and Estimated Value

Table 4b: Summary of Bias for MCMC Parameter Estimates across Items for i=30

Sample size	Order effect condition	α	δ_1	δ_2	δ_3	e	f
500	Small e, small f	.14	.13	.05	.04	.02	.02
	Small e, large f	.15	.13	.06	.03	.02	.03
	Large e, small f	.14	.15	.06	.02	.02	.02
	Large e, Large e	.15	.13	.07	.03	.02	.02
1000	Small e, small f	.08	.08	.05	.01	.01	.02
	Small e, large f	.08	.08	.04	.001	.02	.01
	Large e, small f	.08	.08	.05	.02	.01	.01
	Large e, Large e	.08	.07	.03	.02	.01	.01
2000	Small e, small f	.04	.05	.03	.01	.002	.003
	Small e, large f	.05	.04	.03	.01	.003	.003
	Large e, small f	.05	.03	.04	.001	.002	.001
	Large e, Large e	.04	.04	.02	.001	.001	.003

Note: Bias = Difference Between True and Estimated Value

For α, estimated means approximated true values for each item rather well (see Tables E1a through E1h where E stands for "Appendix E"). In particular, the direction of change for the true versus estimated values were the same across items (i.e., as the true value increased or decreased, the estimated mean increased or decreased, respectively), though the difference between estimated and true values varied substantially across items. Error measurement values, in general, varied substantially across items within each condition, but they decreased across sample size for both the i=15 and i=30 conditions (see Tables 1a through 3b). More specifically, however, for the i=15 conditions, the average decrease in SD and RMSE (across items) seemed to be greatest between n=500 and n=1,000 conditions, whereas the changes in values across samples sizes were more even for MCMC SE. For test length (i=15 to i=30), however, error measurements did not consistently decrease as the number of items in the test increased. As will be seen, some measurements of error for α are larger compared to other parameters, while other error measurement values are smaller, but many were comparable to one another.

For δ_1, parameter recovery was relatively good; namely, the direction of change in estimated parameters mirrored those of the true parameters, though differences between estimated and true values varied quite a bit across items (see Tables E2a through E2h). In addition, summary values for these error measurements are located in Tables 1a through 3b. As can be seen in these tables, on average, error measurement values decreased across sample size. These changes were relatively even for SD and MCMC SE but greater between n=500 and n=1000 for i=30 conditions. Although there were occasional decreases in error measurement for test length (from i=15 to i=30) for some items, there were no overall (average) decreases noted.

Moreover, there was variation in parameter recovery effectiveness compared to other parameters. Some error measurements were comparable to, improved, or worsened relative to α. These differences, however, were very small – i.e., just a few hundredths of a point. For i=15 conditions, SDs were approximately the same for α and δ_1 across all sample sizes. SDs were also

comparable between the two parameters when sample sizes were 1000 and 2000, but the SD values for δ_1 were smaller than that of α for n=500 only (.12 to .14 versus .09-.12). δ_1 MCMC SE values were higher than those of α for all sample sizes (.15, .11, and .08-.09 versus .12-.14, .07-.09, and .05-.06 for n=500, 1000, and 2000, respectively). The RMSEs for δ_1 were higher than that of α for n=500 only (.16-.19 versus .13-.16). For i=30 conditions, SDs were slightly larger for δ_1 than α for all sample sizes (.13-.14, .08-.10, and .05-.08 versus .10-.11, .07-.09, and .04-.06 for n=500, 1000, and 2000, respectively). The MCMC SEs for δ_1 were, on average, slightly higher than that of α for n=500 only (.10-.15 versus .11-.12). Again, for RMSE, values for δ_1 were, on average, slightly higher than that of α for n=500 only (.05-.09 versus .05-.07).

For δ_2, the parameter recovery was adequate; namely, although the discrepancies between true and estimated values varied quite a bit across items, the directions of change for mean estimated values were consistent with true values. Error measurement values, in general, varied substantially across items within each condition (see Tables E3a through E3h), but they decreased across sample size for both the i=15 and i=30 conditions (see summaries of error values in Tables 1a through 3b). More specifically, however, for the i=15 conditions, the average decrease in SD and RMSE (across items) seemed to be greatest between n=500 and n=1,000 conditions, whereas the changes in values across samples sizes were more even for MCMC SE. For test length (i=15 to i=30), however, error measurements did not consistently decrease as the number of items in the test increased.

When comparing parameter recovery between δ_2 and δ_1 or δ_2 and α, the discrepancies were mostly small (a few hundredths of a point). For i=15 conditions, δ_2 compared to δ_1 as follows. All error measurement values were comparable for n=2000. However, for n=500, SDs, MCMC SEs, and RMSEs were lower for δ_2 than δ_1 by a few hundredths of a point (.09-.12 versus .12-.14, .09-.11 versus .12-.14, and .10-.13 versus .07-.08, respectively). Moreover, RMSEs were lower for δ_2 than δ_1 for n=1000 (.07-.08 versus .09-.10). Parameter recovery for δ_2 compared to

α was as follows. For SD, δ_2 values were slightly larger than that of α for n=500 only (.09-.12 versus .12-.14). For MCMC SE, δ_2 values were slightly larger than that of α for all sample sizes (.12-.14, .08-.09, and .05-.06 versus .15, .11, and .08-.09, respectively). RMSE values did not differ noticeably between δ_2 and α for any sample size conditions.

For i=30 conditions, δ_2 compared to δ_1 as follows. All error measurement values were comparable for n=2000. However, for n=500, SDs and RMSEs were lower for δ_2 than δ_1 by a few hundredths of a point (.09-.11 versus .13-.14 and .10-.13 versus .19, respectively). Moreover, RMSEs were lower for δ_2 than δ_1 for n=1000 (.08-.10 versus .10-.12). Parameter recovery for δ_2 compared to α was as follows. There were no discrepancies for SDs or MCMC SEs. However, RMSE δ_2 values were lower than that of α for n=500 and n=1000 (.10-.13 and .08-.10 versus .17-.18 and .10-.12) but not n=2000.

For the final item parameter, δ_3, recovery was adequate. Namely, although the differences between mean estimated versus true values varied across items, these two sets of value changed in the same direction as one another. In general, there was variance across items in the effectiveness of parameter recovery as measured by all types of error (see Tables E4a through E4h). Nevertheless, all error measurements, on average, decreased as sample size increased, though this was not the case – for any error measurement – when taking into account test length (see Tables 1a through 3b). For SDs, there was a greater decrease in value between n=500 and n=1000 for i=30, but change in SD across sample sizes was even for i=15. For MCMC SEs and RMSEs, decreases in values across sample sizes were even for all test lengths.

When comparing parameter recovery between δ_3 and δ_2, δ_3 and $\delta_1$or δ_3 and α, the discrepancies were mostly small (a few hundredths of a point). For i=15 conditions, all error measurement values for δ_3 were, on average, approximate to those of δ_2. Parameter recovery for δ_3 compared to δ_1 was as follows. For all three measurements of error, average values for

δ_3 were lower than those of δ_1 for n=500 (.11 versus .12-.14, .10-.12 versus .12-.14, and .11-.12 versus .16-.19 for SD, MCMC SE, and RMSE, respectively). RMSE values for δ_3 were slightly lower than those of δ_1 for n=1000 and n=2000 as well (.07-.09 versus .09-.10 and .04-.06 versus .06-.07, respectively). Parameter recovery for δ_3 was generally better than that of α. For SD, values were, on average, lower for δ_3 than those of α for n=500 only (.07-.09 versus .12-.14). MCMC SE values for δ_3 were, on average, lower than those of α for all sample sizes (.10-.12 versus .15, .07-.08 versus.11, and .04-.06 versus .08-.09 for n=500, 1000, and 2000, respectively). Finally, RMSE values for δ_3 were, on average, lower than those of α for n=500 and n=2000 (.11-.12 versus .13-.16 and .04-.06 versus .06-.09, respectively).

For i=30 conditions, MCMC SE and RMSE values for δ_3 were, on average, approximate to those of δ_2. However, for SD, δ_3 values were higher than that of δ_2 for n=500 only (.11-.14 versus .09-.11). δ_3 compared to δ_1 as follows. SD and MCMC SE values were, on average, comparable between the two step parameters for all sample sizes. For RMSE, δ_3 values were, on average, lower than those of δ_1 for n=500 and n=1000 (.11-.14 versus 19 and .08-.09 versus .10-.12, respectively). Parameter recovery for δ_3 compared to α was as follows. MCMC SE values were not noticeably different between the two parameters. SD values for δ_3 were, on average, higher than those of α for n=500 only (.11-.14 versus .10-.11). RMSE values for δ_3 were, on average, lower than those of α for n=500 and n=1000 (.11-.14 versus .17-.18 and .08-.09 versus .10-.12, respectively).

In terms of the recovery of the *e* and *f* parameters, the SDs, SEs, RMSEs, and biases across conditions can be found in Tables 1a through 4b. As can be seen, these values were recovered quite well with relatively small error measurements compared to other model parameters. Moreover, all measurements of error decrease as both sample size increases.

Although there was substantial variation among items in the value magnitude for each error measure (see Appendix E), the parameters were, in general, adequately recovered in the simulation study (see Tables 1a through 4b). However, for all parameters, there was a general decrease in all measurements of error as sample size increased – the greatest occurring between n=500 and n=1000 for some measurements and parameters. The same pattern was not observed, however, for test size. Tables 4a and 4b contain average bias in parameter estimation for all conditions, and it can be seen in these tables that changes in bias value is somewhat consistent with other parameters (i.e., decrease as sample size increases, but no such decrease as test size increases). However, for some parameters, this pattern of change is not observed. Thus, it appears that a majority of the change in RMSE across sample size for these latter parameters is attributable to decreases in SD rather than true versus estimated parameters (i.e., bias). In terms of comparison of recovery among parameters, most discrepancies in error measurements were small – i.e., only a few hundredths of a point. Moreover, discrepancies, when present, were usually for sample sizes of n=500, regardless of test length. However, there was no consistent pattern in which parameters were best recovered (i.e., magnitude of values) for any given measurement of error.

To further elaborate upon recovery effectiveness, 95% confidence intervals were constructed around the mean estimates for each parameter within each condition in order to determine the extent to which the estimates match the true values as well as how well these estimates are contained within the confidence intervals. These confidence intervals were calculated in WinBUGS and are also referred to as the "posterior intervals." Specifically, 2.5% and 97.5% percentiles of the posterior samples for each parameter give a 95% posterior "credible" interval, which is the Bayesian analogue of the 95% confidence interval. However, with a Bayesian analysis, we state that there is a 95% probability that the parameter is between the 2.5% and 97.5% interval values. In a conventional, frequentist analysis, on the other hand, we state that

95% of all such intervals will contain the true, but unknown value for the parameter (assuming that the null hypothesis is correct). It is also worth noting that MCMC SEs (i.e., PSDs) were used in computations of these intervals rather than SDs.

For the aforementioned best and worst conditions (from one sample), plus two additional conditions (n=1000, i=15, large order effects and n=1000, i=30, small order effects), all parameter mean estimates (across cells) were graphed against their respective true values and confidence intervals and can be found in Figures 3a through 3l). As expected, the confidence interval surrounding the mean estimate for each parameter tightened as sample size and test length increased. However, it is worth noting that, for α, the intervals were generally larger relative to those of other parameters, and true values for a few items fell outside the confidence interval, though this finding occurred in the worst case scenario condition (n=500, i=15, large order effects). These results are consistent with the recovery effectiveness previously detailed. A further examination of convergence via trace plots did not yield any abnormalities in the estimation process for this parameter; therefore, this aberrance may be attributed to error associated with the lower sample size and test length.

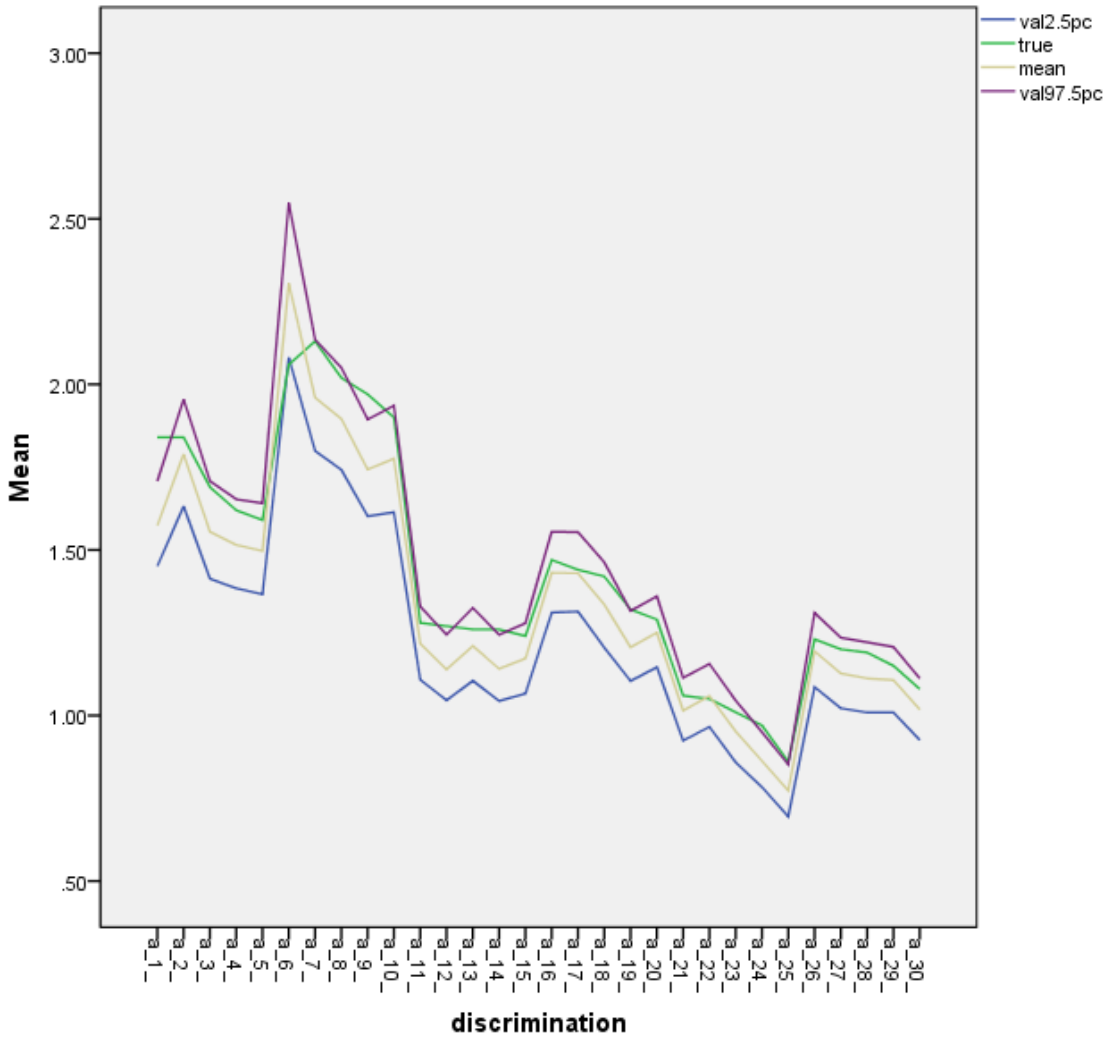


Figure 3a: 95% Confidence Intervals for Item Discrimination, N=2000, I=30, Small e, Small f Condition

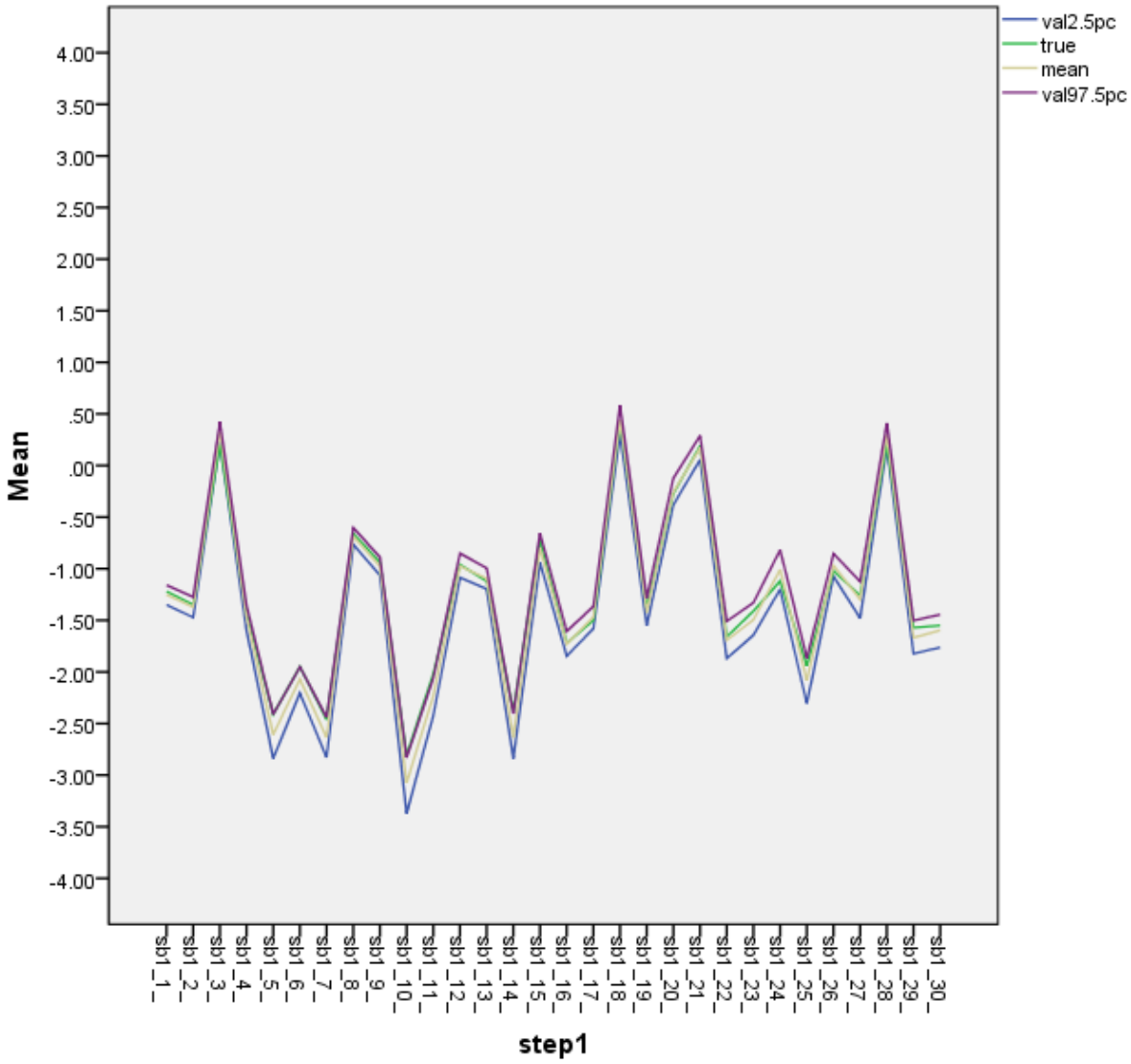


Figure 3b: 95% Confidence Intervals for Step One, N=2000, I=30, Small e, Small f Condition

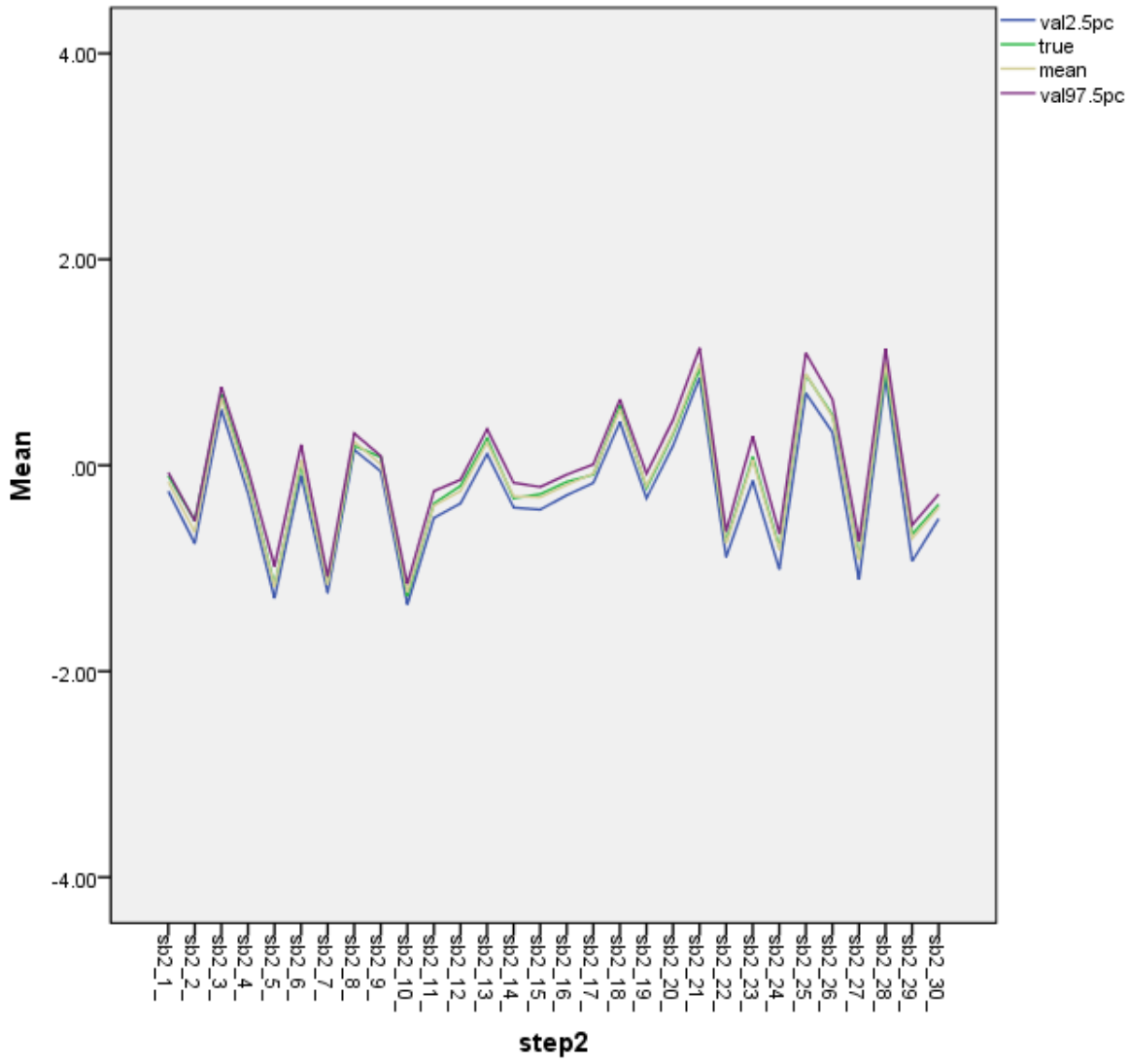


Figure 3c: 95% Confidence Intervals for Step Two, N=2000, I=30, Small e, Small f Condition

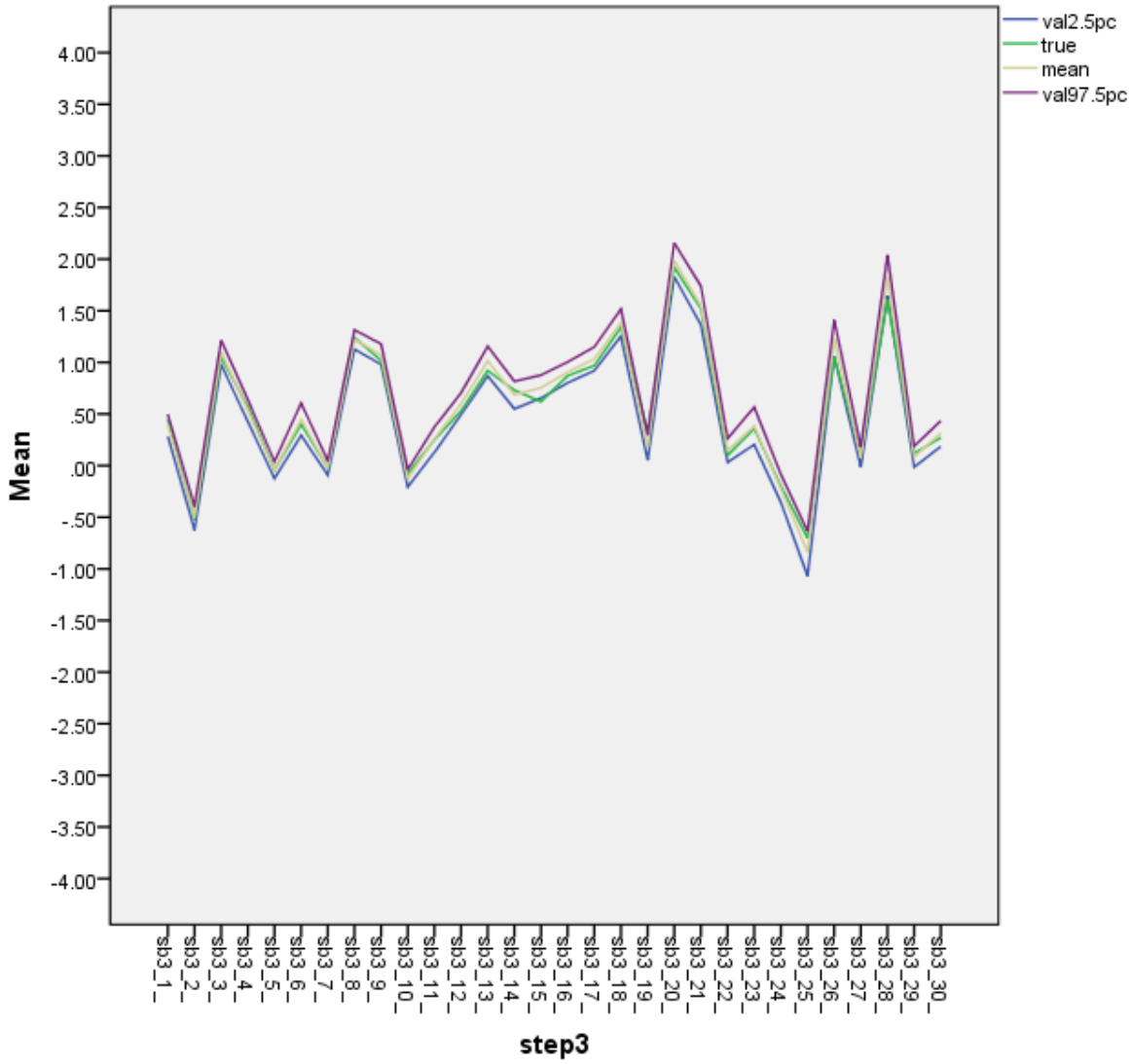


Figure 3d: 95% Confidence Intervals for Step Three, N=2000, I=30, Small e, Small f Condition

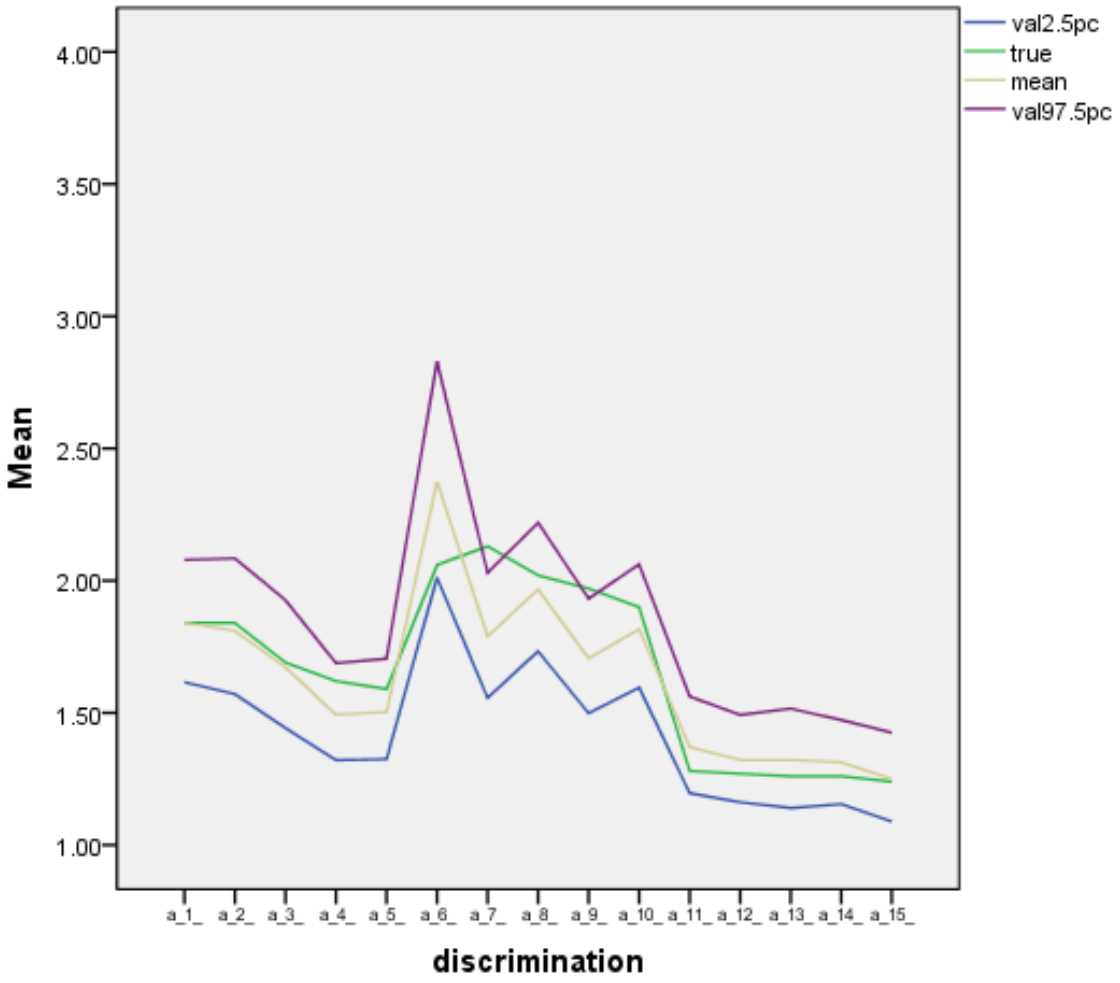


Figure 3e: 95% Confidence Intervals for Item Discrimination, N=1000, I=15, Large e, Large f Condition

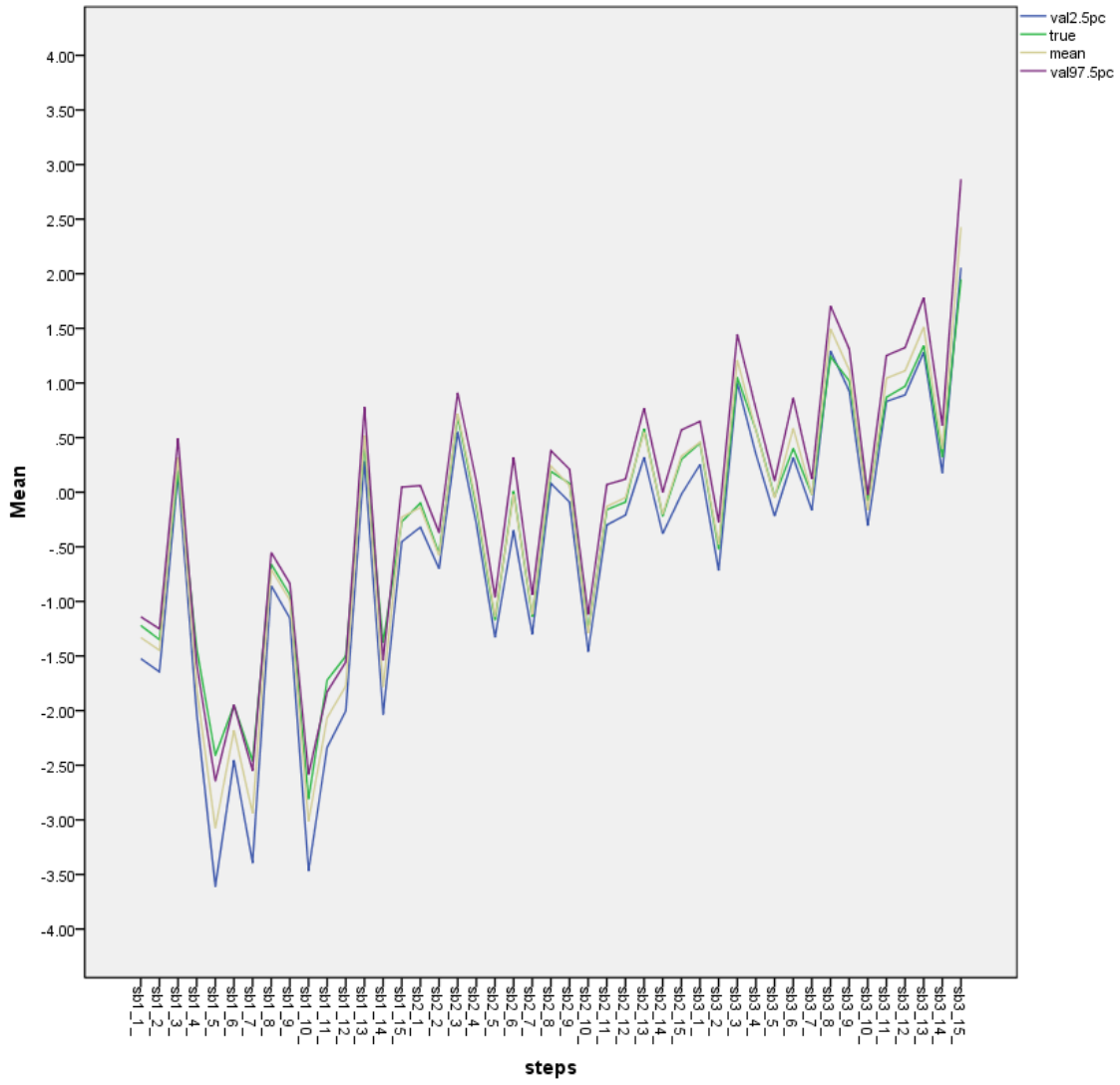


Figure 3f: 95% Confidence Intervals for Step Parameters, N=1000, I=15, Large e, Large f Condition

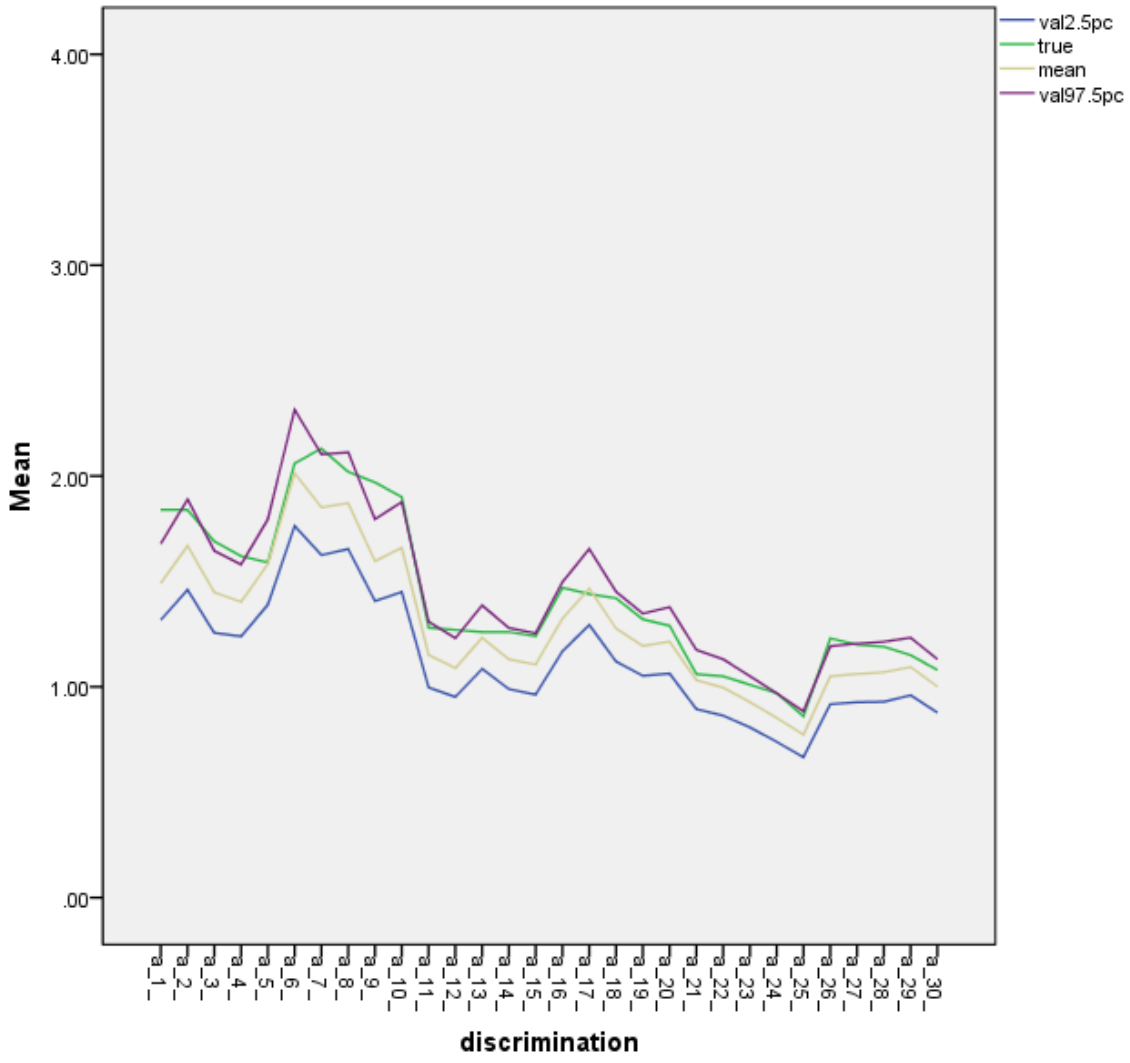


Figure 3g: 95% Confidence Intervals for Item Discrimination, N=1000, I=30, Small e, Small f Condition

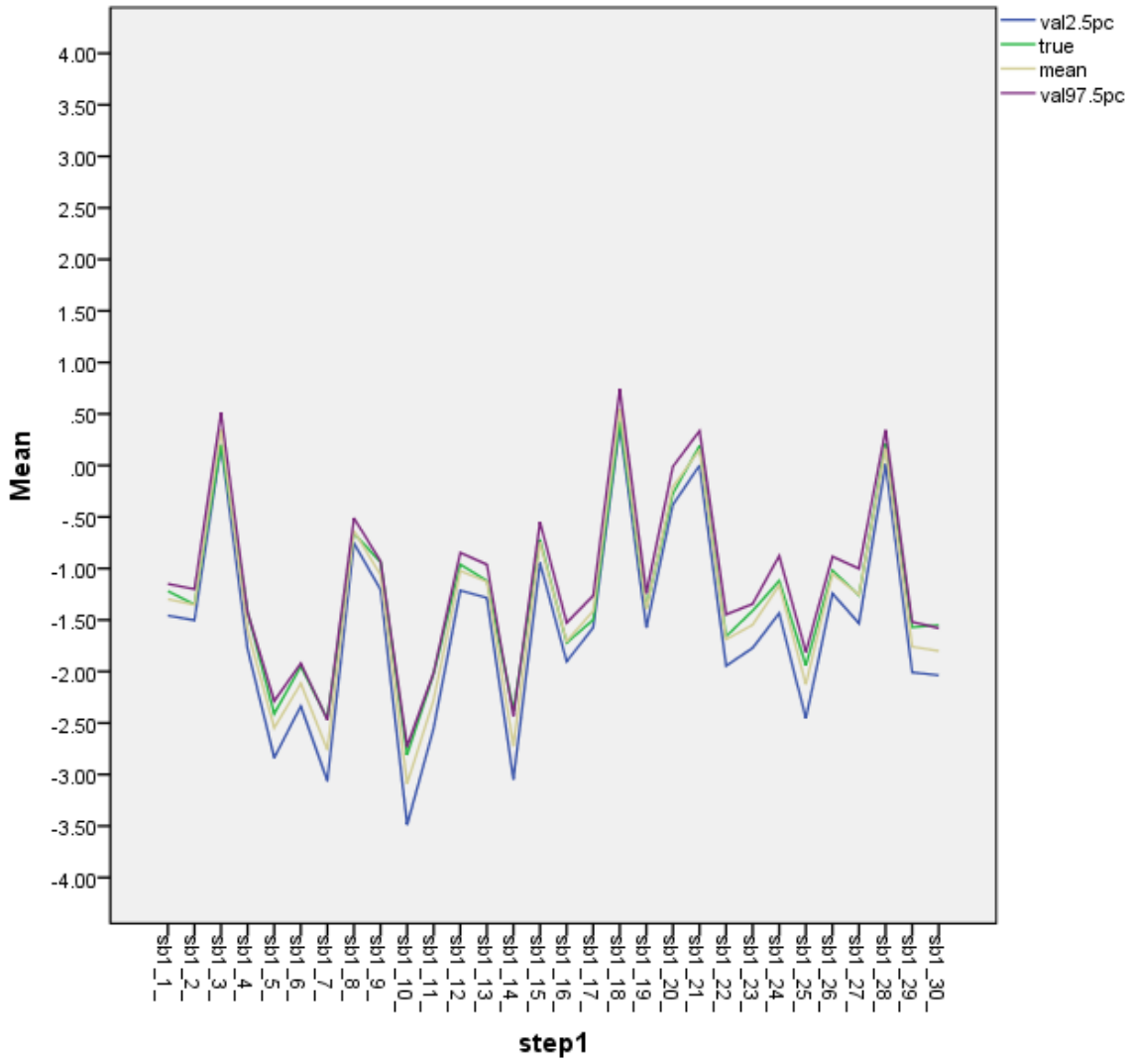


Figure 3h: 95% Confidence Intervals for Step One, N=1000, I=30, Small e, Small f Condition

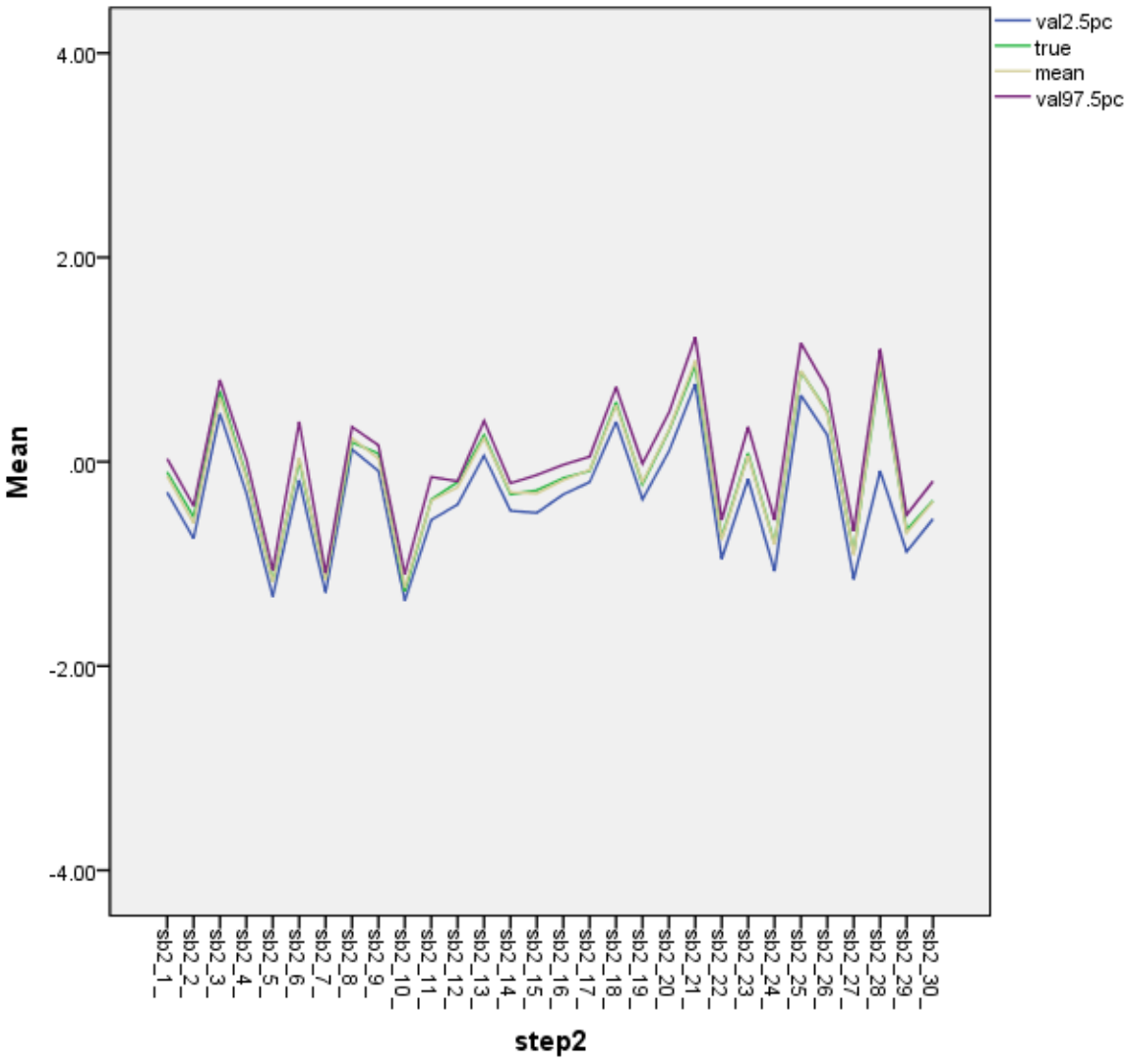


Figure 3i: 95% Confidence Intervals for Step Two, N=1000, I=30, Small e, Small f Condition

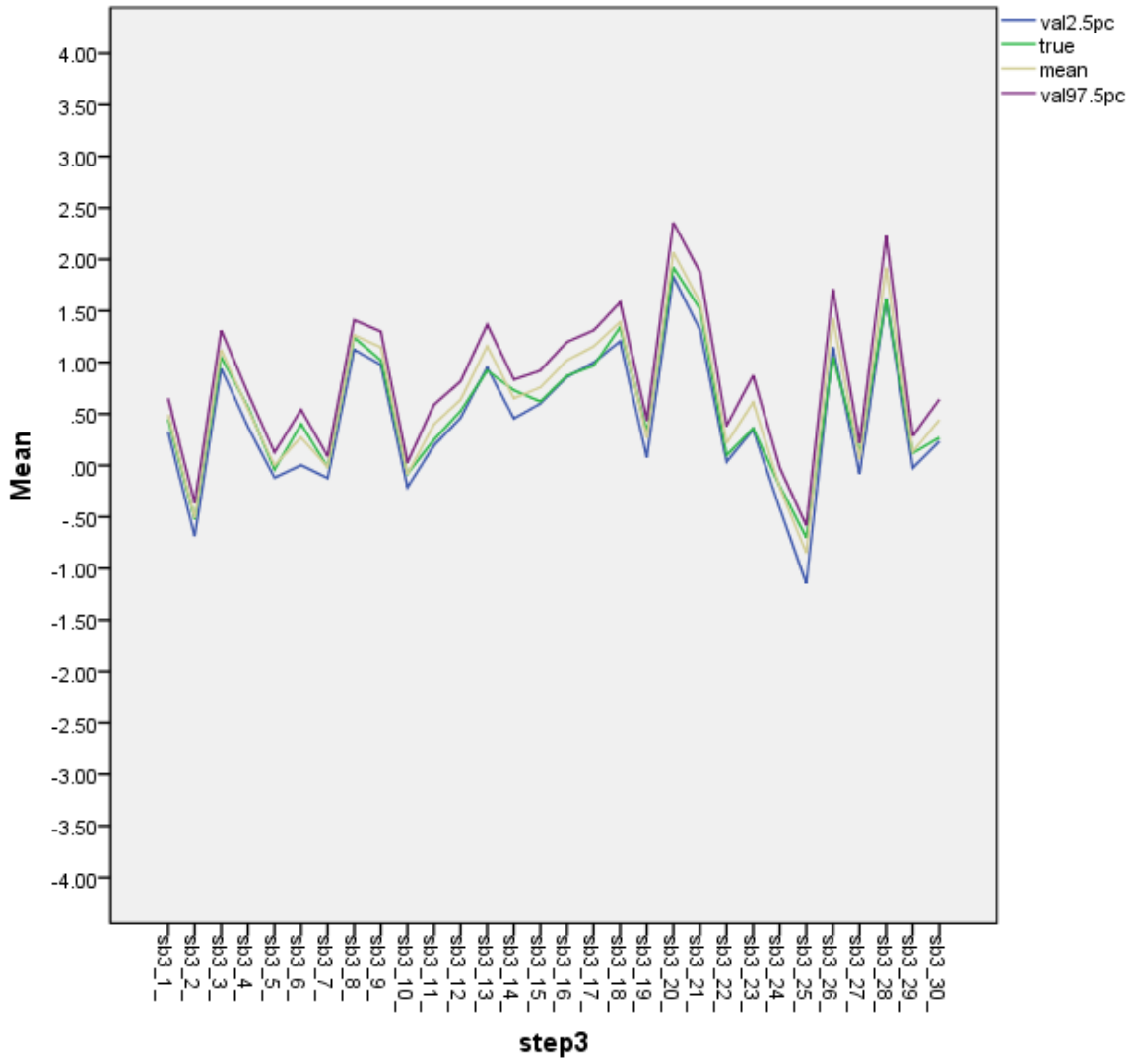


Figure 3j: 95% Confidence Intervals for Step Three, N=1000, I=30, Small e, Small f Condition

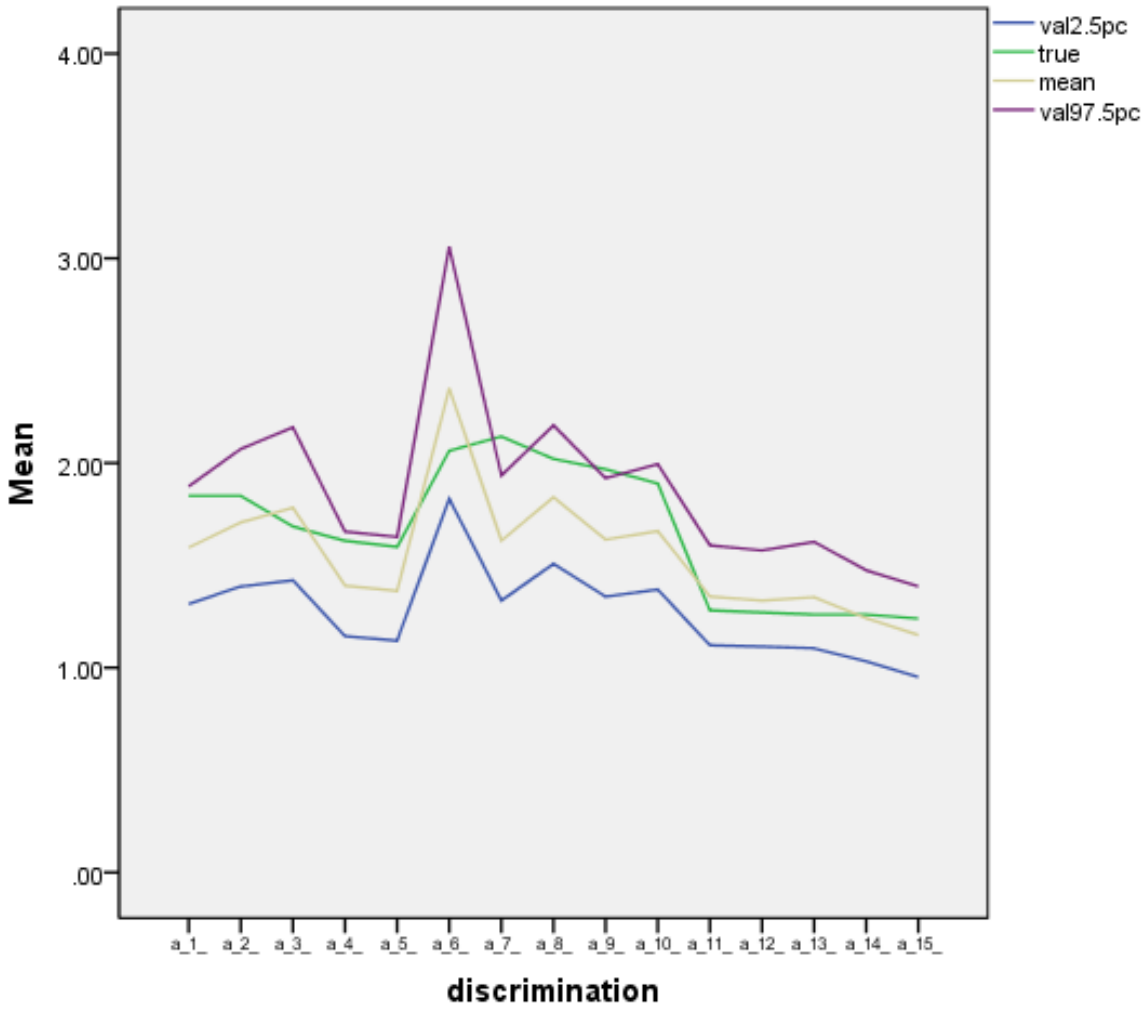


Figure 3k: 95% Confidence Intervals for Item Discrimination, N=500, I=15, Large e, Large f Condition

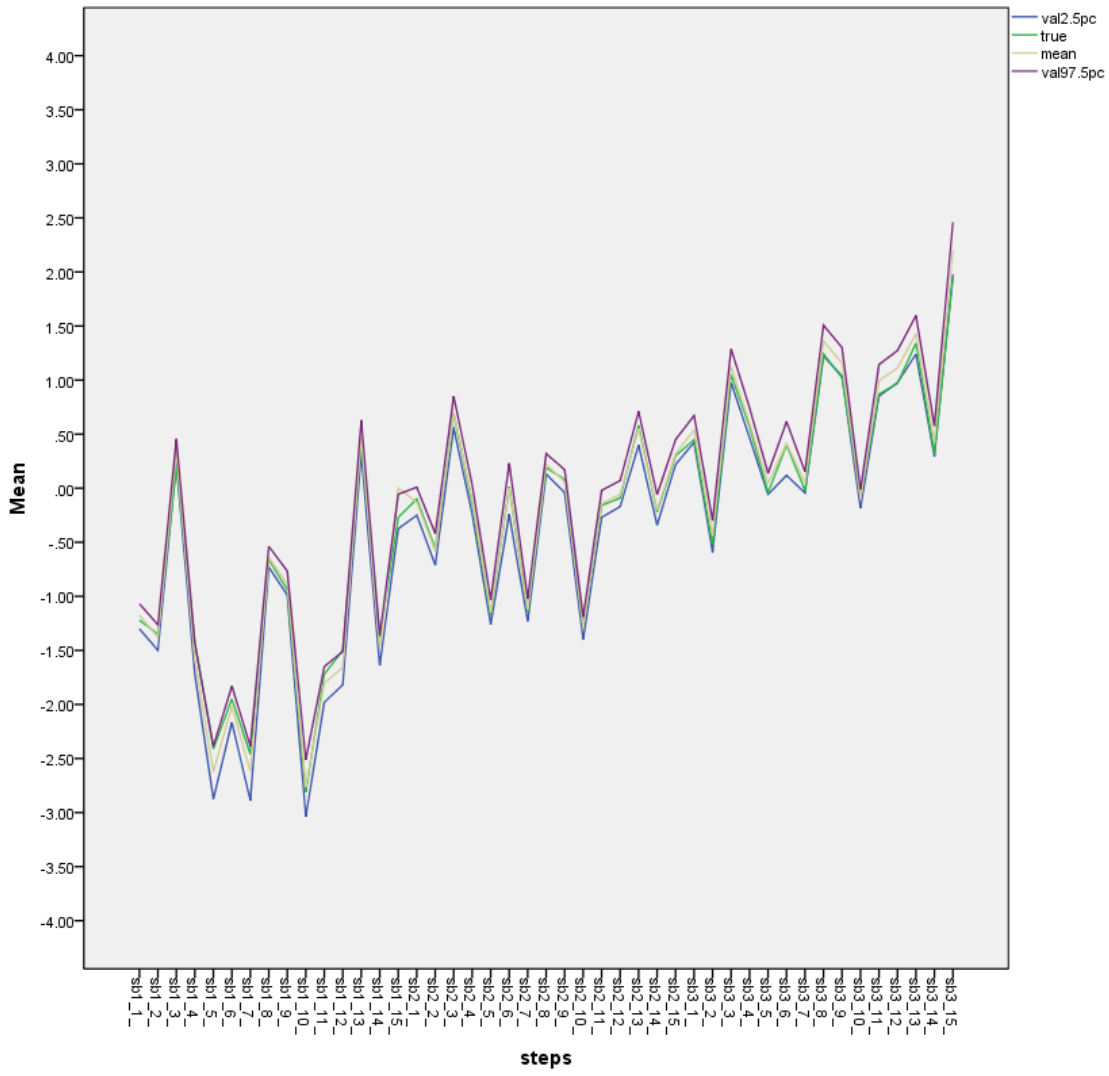


Figure 3l: 95% Confidence Intervals for Step Parameters, N=500, I=15, Large e, Large f Condition

Finally, the correlation between true and estimated trait level values was computed and graphed for the worst and best case conditions (see Figures 4a and 4b, respectively). As expected, the correlation increased in the best case condition, indicating that recovery was adequate but varied as a function of at least sample size or test length.

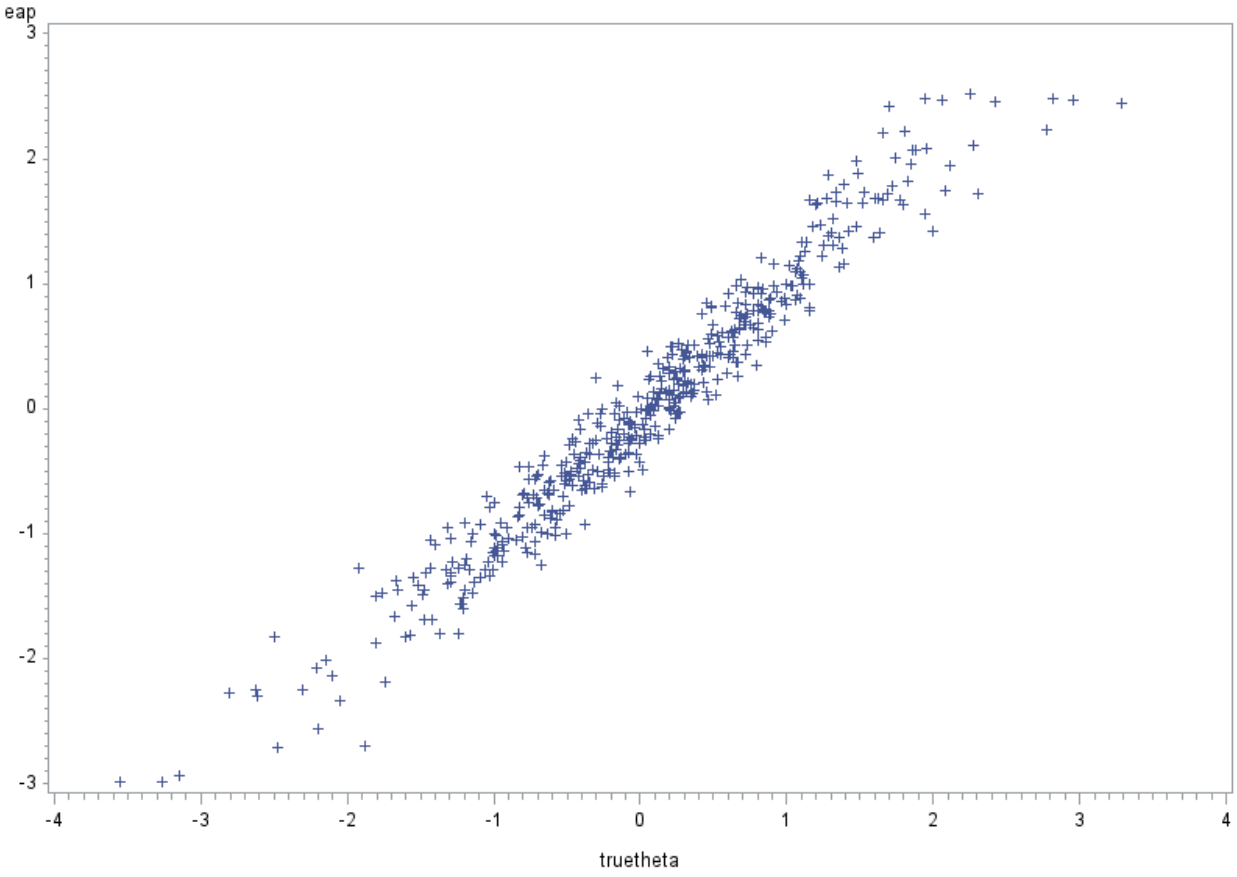


Figure 4a: Correlation between True and Estimated Latent Trait Values for N=500, I=15, Large E and Large F Condition

$r = .975$ ($p < .0001$)

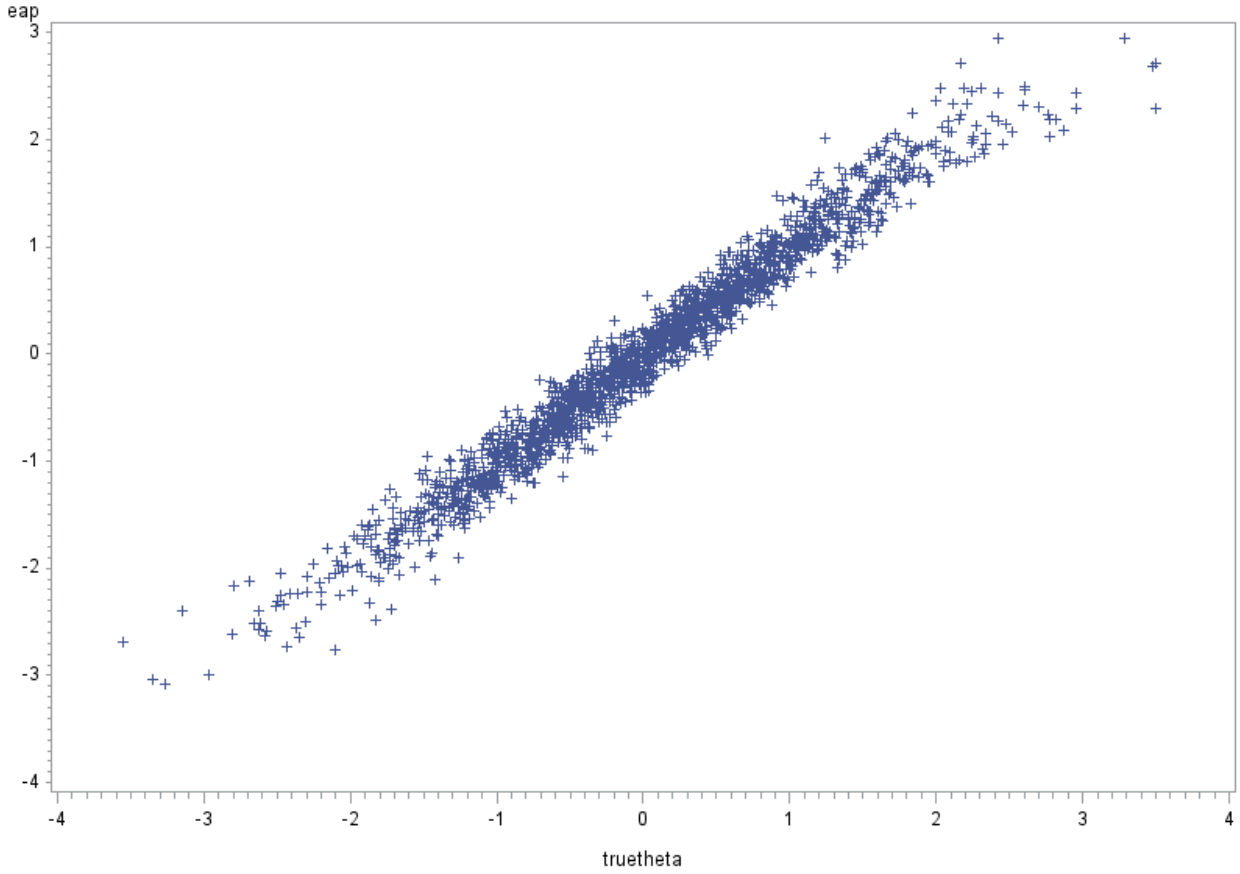


Figure 4b: Correlation between True and Estimated Latent Trait Values for N=2000, I=30, Small E and Small F Condition

$r = .982$ ($p < .0001$)

Second, an ANOVA was performed based on the aforementioned design (3 x 2 x 2 x 2): sample size, number of items, the pattern of values for the order-step parameters (*f*), and the pattern of values for the order-discrimination parameters (*e*), where SD, MCMC SE, and RMSE were the dependent variables (again, overview of means for these values are located in Tables 1a through 3b). The purpose of including the order effect parameters, *f* and *e*, was to see how their magnitudes impact the accuracy of the remaining model parameter estimates. For example, it was expected that large *e* values would impact the accuracy of θ since this *e* is interpreted as a factor impacting discrimination, i.e., the precision of latent trait measurement as described within an IRT model. *e* parameters, on the other hand, were expected to impact, if anything, other item parameters such as item steps. However, it may be that, although they impact step parameters (δ), the magnitude of *f* does not affect the accuracy of these parameter estimates. In order to control for Type I error, given that the ANOVA will be conducted five times – once for each of the five model parameters (not including *e* and *f*, which are part of the design, itself), α was set to .05/5=.01 for each test.

As expected, for all parameters, all three types of error decreased as sample size increased. For RMSE, *F*-values in which $p<.01$ were $F(2, 719)=87.750$ for θ, $F(2, 719)=32.020$ for α, $F(2, 719)=202.075$ for δ_1, $F(2, 719)=422.021$ for δ_2, and $F(2, 719)=378.005$ for δ_3. With respect to MCMC SE, the *F*-test values in which $p<.01$ were as follows. For θ, $F(2, 719)=6.994$, for α, $F(2, 719)=15.761$, for δ_1, $F(2, 719)=321.175$, for δ_2 , $F(2, 719)=201.525$, and for δ_3, $F(2, 719)=362.025$. Finally, for SDs, the *F*-test values with $p<.01$ were as follows. For θ, $F(2, 719)=71.001$, for α, $F(2, 719)=34.992$, for δ_1, $F(2, 719)=287.352$, for δ_2 , $F(2, 719)=155.234$, and for δ_3, $F(2, 719)=192.998$. Therefore, large sample sizes (i.e., above 500) are ideal for obtaining accurate, stable estimates for all item parameters of the GPCFM.

An increase in test size from 15 to 30 items resulted in a statistically significant decrease in all error measurements for θ only. For RMSE, the *F*-value with $p<.01$ was $F(1, 719)=99.080$. For

MCMC SE, the *F*-test value with $p<.01$ was $F(1, 719)=39.048$. For SD, the *F*-test value was $F(1, 719)=75.997$ ($p<.01$). Thus, when seeking accurate estimates of a person's standing on a latent trait, it is, in general, ideal to use the GPCFM when you have more than 15 items in a test.

With respect to the order effect parameters, *e* and *f*, the following results were found. As *e* increased, there was a statistically significant decrease in RMSE, MCMC SE, and SD of θ [$F(2, 719)=29.001$ ($p < .01$), $F(2, 719)=30.963$ ($p < .01$) and $F(2, 719)=34.579$ ($p < .01$), respectively] but no main effects of change in error in the estimates of item parameters. There were no noticeable changes in any error measurements of item or person parameter estimates as a function of *f* magnitude. Thus, as expected, an increase in the order effect in terms of its impact on α is most advantageous for decreasing error in θ estimates. This is logical when one considers that α is a reflection of an item's ability to adequately measure a respondent's θ.

Only one interaction effect was found to be statistically significant. Specifically, an increase in both sample size and test length resulted in greater accuracy for α on all measures of error[$F(2, 719)=19.002$ ($p < .01$) for RMSE, $F(2, 719)=9.659$ ($p < .01$) for MCMC error, and $F(2, 719)=5.467$ ($p < .01$) for SD]. Thus, larger sample sizes and test lengths are recommended for accurate measures of item α,.

Finally, correlations between true and estimated theta for one replication from conditions representing the best and worst-case scenarios are depicted in Figures 4a and 4b. All correlations were in the upper .90s. Thus, recovery of theta was deemed good. In general, parameters were well recovered and, thus, the model appears to be tenable in theory. Therefore, further analysis on real data was conducted.

Application to Real Data

Participants and Procedure

One real data set was used to demonstrate the measurement of order effects. Eight hundred and seventeen male adult respondents were sampled from a military base. The sample completed a test containing an early version of the big five via computer. The order of items was randomized for each participant. Total test time was 90 minutes.

Personality Instrument

A set of items that represent an early version of the Big Five was used in the current study (Christal, 1993). Thus, five traits were measured in the real data sample. The total number of items was 30 for conscientiousness, 38 for neuroticism, 35 for agreeableness, 31 for extroversion, and 29 for openness. The response format was a 45-point scale, ranging from -22 to +22 ("completely unlike me" to "perfectly describes me"). These scores were polytomized into a five-point scale similar to the Big Five NEO-PI-R (Costa & McCrae, 1992). A sample of items is included in Appendix E.

Preliminary Analysis

Before the model was applied, the scores on conscientiousness was submitted to a principle components analysis, PCA, (based on polychoric correlations) in order to check the dimensionality of the items. Unidimensionality in the data was confirmed in several ways. First, the number of components with eigenvalues greater than 1 was identified. Next, the eigenvalue plot was examined in order to determine if there was a large drop from the first to second factor. Moreover, the amount of the variance in item scores attributable to the first component was be examined.

For the data on each of the five traits, unidimensionality was supported. Namely, few factors had eigenvalues above 1, there was a substantial drop in the eigenvalue plot between the first and second factors, and at least 60% of the variance in total item scores was attributable to the first factor. Thus, the data for all five traits were deemed unidimensional and subsequent analysis were conducted.

Model Fit

Next, model fit was examined by looking at the SDs (MCMC SEs) and Monte Carlo (MC) errors of the estimates for each parameter in the model. These error measures for each of the seven model parameters for each of the five traits are located in Tables 5a through 5e (one table per trait), where the mean estimated value was referred to as "Estimate" and the SD was referred to as "Standard Error." SD values were comparable to those of MCMC SEs uncovered in the simulations. It is also worth noting that there were 1,000 burn-in samples and 20,000 subsequent iteration values retained (wherein every 5^{th} iteration value was retained) in order to compute the posterior means, SDs, and MC errors. Justification for these latter steps is based on the simulation study in which similar methodology resulted in model convergence. Taken together, the error measurements indicate adequate model fit as well as a high likelihood of model convergence.

Next, a comparison of fit among three nested models was conducted. Namely, the new model, the GPCFM, was compared to two models nested within the GPCFM – namely, the GPCM and the GFM – in order to determine which model best fits the data (see Table 5). The statistical criterion used to judge this comparison was the Deviance Information Criterion (DIC; Spiegelhalter et al., 2002). The DIC index functions similarly to the Akaike Information Criterion (AIC; Akaike, 1973) and the Bayesian Information Criterion (BIC; Schwartz, 1978) because it weighs model complexity (number of parameters) with information (sample size, etc.). As with the AIC and BIC, the lower the index, the better the model fit. As can be seen in Table 5, the

GPCFM provided the best fit to the data for neuroticism, openness, and agreeableness. There was some degree of improvement in the DIC for conscientiousness, but it was rather small in comparison. These results are consistent with the findings regarding the *e* magnitudes and, hence, the presence of order effects previously discussed.

Table 5: Comparison of Model Fit

Trait	*Model*	*DIC*
Neuroticism	**GPCM**	**58,770.00**
	GFM	**58,630.00**
	GPCFM	**58,550.00**
Extroversion	**GPCM**	**62,490.00**
	GFM	**62,420.00**
	GPCFM	**62,420.00**
Openness	**GPCM**	**59,320.00**
	GFM	**59,240.00**
	GPCFM	**59,150.00**
Agreeableness	**GPCM**	**48,330.00**
	GFM	**48,230.00**
	GPCFM	**48,100.00**
Conscientiousness	**GPCM**	**54,530.00**
	GFM	**54,420.00**
	GPCFM	**54,390.00**

***N*=817**

Results

In terms of the *e* and *f* sizes uncovered from these data sets, the results were somewhat unexpected (see Tables 6a through 6e). For example, the *f* parameter values were higher than expected in several traits (neuroticism, agreeableness, openness, and conscientiousness), whereas the *e* parameter values were lower than expected in all traits except, perhaps, agreeableness, where *e* reached moderate levels and increased across blocks. Specifically, the results of the real data analysis were as follows. The type and size of the order effect (*e* or *f*) was detected and differed for most traits. Namely, a low *f* in block 2 but a high *f* in block 3 was observed for agreeableness. At the same time, *e*s for both blocks 2 and 3 were moderate for agreeableness. For neuroticism, high *f*s were observed for both blocks, each block value increasing relative to the previous block value, and low *e*s were observed for both blocks, though these value increased across block. Similar results were found for openness. Finally, conscientiousness yielded low to non-existent values of *e*, but these values did increase across blocks. Nevertheless, conscientiousness demonstrated the highest *f* values, which increased across blocks. Thus, an increase in order effects – both *e* and *f* - was, to some degree observed for all traits except extroversion. Specifically, the thresholds decreased over time, resulting in higher levels of agreement with an item statement as the item appeared later in the test. Moreover, items became increasingly discriminating as they appeared later in the test. The presence of the latter effects may be explained by a streamline in the response process (Tourangeau & Rasinski, 1988), but the higher-than-expected order effects, as represented in *f*, suggest the presence of social desirability bias. In fact, the self-schema retrieval process may have been biased toward the retrieval of scant behavioral examples consistent with the socially-desirable traits – i.e., selective examples of when a person is agreeable, open to experience, and more emotionally stable.

Table 6a: Model Fit and MCMC Parameter Estimates for Neuroticism

MODEL PARAMETER	***Estimate***	***Standard Error***	***MCMC Error***
Trait Level			
$\boldsymbol{\theta_p}$	-.97	.19	.003
Item Discrimination			
$\boldsymbol{\alpha_i}$	.98	.05	.001
Category Thresholds (Steps)			
$\boldsymbol{\delta_{i1}}$	-.20	.15	.003
$\boldsymbol{\delta_{i2}}$	.61	.19	.003
$\boldsymbol{\delta_{i3}}$	.92	.18	.003
$\boldsymbol{\delta_{i4}}$	1.22	.16	.003
Order Effect on Category Threshold			
$\boldsymbol{f_2}$	.10	.02	.001
$\boldsymbol{f_3}$	.12	.03	.002
Order Effect on Item Discrimination			
$\boldsymbol{e_2}$	1.00	.03	.001
$\boldsymbol{e_3}$	1.0640	.0427	.001

n=817

Note: Estimate = the estimated posterior mean for a given parameter

Note: Standard Error = Variation of iteration values in a chain (same as MCMC SE from simulation study)

Note: MCMC Error = Estimation error attributable to autocorrelations within the chain (i.e., correlations among sampled or iteration values)

Note: f and e values are not means but single estimates for blocks 2 and 3, respectively

Table 6b: Model Fit and MCMC Parameter Estimates for Extroversion

MODEL PARAMETER	***Estimate***	***Standard Error***	***MCMC Error***
Trait Level			
$\boldsymbol{\theta_p}$	.05	.19	.003
Item Discrimination			
$\boldsymbol{\alpha_i}$	.81	.05	.002
Category Thresholds (Steps)			
$\boldsymbol{\delta_{i1}}$	-1.56	.19	.004
$\boldsymbol{\delta_{i2}}$	-.84	.18	.003
$\boldsymbol{\delta_{i3}}$	.07	.19	.003
$\boldsymbol{\delta_{i4}}$	.25	.18	.003
Order Effect on Category Thresholds (Steps)			
$\boldsymbol{f_2}$	-.06	.02	.001
$\boldsymbol{f_3}$	-.08	.02	.001
Order Effect on Item Discrimination			
$\boldsymbol{e_2}$	1.00	.02	.0004
$\boldsymbol{e_3}$	.97	.02	.0004

n=817

Note: Estimate = the estimated posterior mean for a given parameter

Note: Standard Error = Variation of iteration values in a chain (same as MCMC SE from simulation study)

Note: MCMC Error = Estimation error attributable to autocorrelations within the chain (i.e., correlations among sampled or iteration values)

Note: f and e values are not means but single estimates for blocks 2 and 3, respectively

Table 6c: Model Fit and MCMC Parameter Estimates for Openness

MODEL PARAMETER	***Estimate***	***Standard Error***	***MCMC Error***
Trait Level			
$\boldsymbol{\theta_p}$	.34	.18	.003
Item Discrimination			
$\boldsymbol{\alpha_i}$	.61	.04	.001
Category Thresholds (Steps)			
$\boldsymbol{\delta_{i1}}$	-1.21	.19	.004
$\boldsymbol{\delta_{i2}}$	-1.03	.19	.004
$\boldsymbol{\delta_{i3}}$	-.68	.19	.003
$\boldsymbol{\delta_{i4}}$	.16	.18	.003
Order Effect on Category Thresholds (Steps)			
$\boldsymbol{f_2}$	.13	.02	.001
$\boldsymbol{f_3}$	.24	.02	.001
Order Effect on Item Discrimination			
$\boldsymbol{e_2}$	1.01	.03	.0004
$\boldsymbol{e_3}$	1.04	.03	.001

n=817

Note: Estimate = the estimated posterior mean for a given parameter

Note: Standard Error = Variation of iteration values in a chain (same as MCMC SE from simulation study)

Note: MCMC Error = Estimation error attributable to autocorrelations within the chain (i.e., correlations among sampled or iteration values)

Note: f and e values are not means but single estimates for blocks 2 and 3, respectively

Table 6d: Model Fit and MCMC Parameter Estimates for Agreeableness

MODEL PARAMETER	***Mean***	***SD***	***MCMC Error***
Trait Level			
$\boldsymbol{\theta_p}$	1.10	.19	.003
Item Discrimination			
$\boldsymbol{\alpha_i}$	.96	.06	.001
Category Thresholds (Steps)			
$\boldsymbol{\delta_{i1}}$	-2.06	.22	.01
$\boldsymbol{\delta_{i2}}$	-2.04	.22	.004
$\boldsymbol{\delta_{i3}}$	-1.87	.18	.005
$\boldsymbol{\delta_{i4}}$	-.97	.18	.002
Order Effects on Category Thresholds (Steps)			
$\boldsymbol{f_2}$	.07	.03	.001
$\boldsymbol{f_3}$	.22	.03	.001
Order Effects on Item Discrimination			
$\boldsymbol{e_2}$	1.14	.03	.001
$\boldsymbol{e_3}$	1.23	.04	.001

n=817

Note: Estimate = the estimated posterior mean for a given parameter

Note: Standard Error = Variation of iteration values in a chain (same as MCMC SE from simulation study)

Note: MCMC Error = Estimation error attributable to autocorrelations within the chain (i.e., correlations among sampled or iteration values)

Note: f and e values are not means but single estimates for blocks 2 and 3, respectively

Table 6e: Model Fit and MCMC Parameter Estimates for Conscientiousness

MODEL PARAMETER	***Estimate***	***Standard Error***	***MCMC Error***
Trait Level			
$\boldsymbol{\theta_p}$	1.25	.19	.002
Item Discrimination			
$\boldsymbol{\alpha_i}$	.98	.07	.001
Category Thresholds (Steps)			
$\boldsymbol{\delta_{i1}}$	-2.03	.22	.005
$\boldsymbol{\delta_{i2}}$	-1.89	.18	.002
$\boldsymbol{\delta_{i3}}$	-1.54	.19	.004
$\boldsymbol{\delta_{i4}}$	-.09	.18	.002
Order Effects on Category Thresholds (Steps)			
$\boldsymbol{f_2}$	.20	.18	.002
$\boldsymbol{f_3}$	.22	.03	.001
Order Effects on Item Discrimination			
$\boldsymbol{e_2}$	1.00	.03	.001
$\boldsymbol{e_3}$	1.01	.03	.001

n=817

Note: Estimate = the estimated posterior mean for a given parameter

Note: Standard Error = Variation of iteration values in a chain (same as MCMC SE from simulation study)

Note: MCMC Error = Estimation error attributable to autocorrelations within the chain (i.e., correlations among sampled or iteration values)

Note: f and e values are not means but single estimates for blocks 2 and 3, respectively

For traits that exhibited notable *e* and *f* order effects, perhaps the increase in *e* can be explained by an initially slow process of searching for these infrequent behavioral examples of the socially-desirable trait, which eventually speeds up as the person has a larger repertoire of (perceived) examples in working memory. The remaining trait, particularly extroversion, did not exhibit notable order effects and, if anything, decreased in both self-rated extroversion and precision in latent trait measurement (i.e., effect of order on item discrimination) over time. Explanation for this finding is perplexing and an explanation is unknown or unfounded in the context of the current study.

CHAPTER 6

CONCLUSIONS

In closing, order effects are extant in self-report personality testing (e.g., Knowles, 1988). This pattern of item responses over time manifests itself as statistical change. This phenomenon is not new and has been studied mostly via CTT methodology, in which parameters and inferences are population-specific. The purpose of the current study was to study the order effect via IRT – namely, by constructing a new IRT model that incorporates order effects into the computation of latent trait values for respondents who take self-report personality tests. This additional step toward identifying and incorporate cognitive processing in personality testing was the scientific contribution of this research. Simulation studies and application of the model to real data determined a) the theoretical utility of the model and b) its viability in situations where real data has demands for the inclusion of this effect such as in applied settings (e.g., selection or educational testing).

In the simulation study, sample size had the most widespread effect on error – namely, SD, MCMC SE, and RMSE for all model parameters. An increase in test length was advantageous for θ only. It is also worth noting that as *e* values increased, all types of error in θ decreased. The parameter *f*, however, did not appear to influence error for any model parameters. The only interaction effect that was significant was that of sample size by test length, which impacted all measures of error for α. Namely, as both sample size and test length increased, error in α estimates decreased. However, the *p*-value level was set to .01 to control for Type I error. Therefore, perhaps more simulations per cell would reveal statistically significant differences in error levels for some or all of the model parameters as test length, *e* and *f* increase, as well as more significant interaction effects. Moreover, it is worth noting that recovery of the order effect

parameters, themselves, was quite good (i.e., small error measurements compared to other model parameters)

In terms of the real data analysis, the new model (GPCFM) provided the best fit for most traits, though this improvement in fit was most substantial for neuroticism, openness, and agreeableness. With respect to order effects, neuroticism, openness, agreeableness, and conscientiousness yielded high values for *f*, and these values increased from block 2 to block 3 for all four traits (see Tables 6a, 6c, 6d, and 6e). These results imply that respondent's self-ratings increased throughout the test for these four traits. On the other hand, mean self-ratings did not increase noticeably, from block to block, for extroversion. In fact, self-ratings appeared to decrease from block to block for extroversion. An upward shift in mean ratings suggests that respondents may be engaging in some form of social desirability bias, whether intentional or unintentional. In fact these order effects were strongest in conscientiousness, as one might expect given the importance of dutifulness and orderliness as crucial aspects of being successful in a military setting. Indeed, the mean θ levels for this trait were higher than those of other traits. Neuroticism was reverse-scored, so it appeared that respondents became increasingly aware that they were being evaluated on this trait and gave increasingly higher ratings, perhaps so as to not appear mentally unstable. A similar process appears likely for agreeableness – as respondents become increasingly aware of being evaluated on this trait, they give themselves higher ratings so as to appear more agreeable and friendly. Perhaps this increase in order effect on step parameters is due to a heightened awareness, on the part of the respondent, of the importance of getting along with and working together with other military personnel in, for example, team-oriented scenarios. There is no explanation, however, for the reverse effect in extroversion.

In terms of *e* parameter values (order effects on item discrimination), only neuroticism, agreeableness, and openness yielded notable values, and these were all small to moderate in magnitude. Therefore, self-ratings became increasingly stable and consistent for these three traits,

over time. However, such a finding did not emerge for extroversion (and was minimal, yet increasing across blocks, for conscientiousness). In fact, the order effect decreased across blocks for extroversion, a finding which is difficult to interpret. Nevertheless, there are a few potential explanations for the preceding results. First of all, perhaps the response process did, as hypothesized, become increasingly streamlined as the respondents were confronted with more and more items measuring the traits neuroticism, openness, and agreeableness. For example, perhaps respondents initially required more time to interpret the content in these items because they do not, otherwise, think about such traits (i.e., in their daily life) or because the respondents do not identify with these traits (i.e., as crucial aspects of their perceived personalities) and had to search longer for examples of when he or she exhibited behaviors consistent with these traits. Yet, as time progressed, the respondents became increasingly aware of the content and were able to more quickly retrieve, from memory, examples of when he or she exhibited behavior consistent with these traits. Alternatively, it is possible that, for extroversion and conscientiousness, respondents did not require as long a period of time to adjust their response process because they were already primed and highly cognizant of these traits. This explanation seems likely given an emphasis in the military on conscientiousness – particularly, dutifulness and orderliness – because these are potentially the most important traits for military personnel. With respect to extroversion, perhaps respondents do not want to appear withdrawn because of an emphasis on group activities and teamwork in a military setting.

It is interesting that most of the traits that demonstrated order effects impacting item step parameters also yielded order effects that influence item discrimination. An exception is conscientiousness, where only the former type of order effect emerged. Perhaps respondents engaged in a more complicated response process for these traits – one in which content interpretation and self-schema trait retrieval were heavily influenced by what the respondent considered socially appropriate (i.e., in a military setting). For example, a respondent may not

identify with certain traits, initially, but wants to present him or herself in a socially desirable manner; hence, the respondent selectively retrieves the few instances in which he or she behaves in such a manner, a process which may take longer in terms of attaining consistency over time. Alternatively, the respondent may knowingly or unknowingly distort self-perception as the test continues (Feldman & Lynch, 1988), thus developing or modifying a self-schema to fit one that is more socially desirable.

A limitation in the current application is that only three blocks of times are considered, rather than each individual position (e.g., 30 if there are 30 items). Therefore, the exact nature of the relationship (e.g., linear, curvilinear) cannot be known. Moreover, the real data analysis is based on a highly specific population: male military personnel. Therefore, future research should involve application of the GPCFM to a wider range of populations. Finally, all five traits were randomized and given to the respondents in one test session. Thus, the test, itself, was multidimensional even though the sets of items for each trait were unidimensional. A potential implication of this multidimensionality within the overall test may have limited the magnitude of the order effects uncovered in the current study. Further research should consider measuring only one trait in a given session, or investigating the usefulness of multidimensional models to account for any traits that covary in a respondent's set of self-schemas. For example, a respondent may consider agreeableness and extroversion to be highly related and a dominant aspect of his or her personality. As a result, it may be observed that responses to items measuring these two traits increase in correlation (with one another) over time rather than the observation of an increase in correlations among scores for items reflecting one, single trait. Also, there are alternative explanations for why both types of order effects occurred within the same trait – namely, a respondent may respond more consistently to items measuring a particular trait not necessarily because he or she is able to more quickly retrieve examples from a true or legitimate self-schema relating to that trait but, rather, an ideal self-schema or one that he or she is devising around the

trait as he or she encounters more and more items (Feldman & Lynch, 1988). The latter seems particularly likely if a) the respondent does not actually identify with the trait as being part of his or her personality or self-schema and/or b) the trait is socially desirable and the respondent has difficulty remembering examples of when he or she behaved in a manner consistent with the trait.

Nevertheless, a few potentially important implications of this order effects model that may useful for future research is that a) scores, by reflecting cognitive process, take into account order effects that might otherwise, if unknown or unaccounted for, compromise the substantive validity of score use (Messick, 1989) and b) there may be respondents, or groups or respondents, who differ in the number of items required to attain stable estimates of trait level. Multiple implications may be derived from these findings. First of all, perhaps the first set of items that represent a respondent's adaptation to the response process, may be discarded to improve the validity of test or subsequent item scores. However, if the latter item scores reflect social desirability bias, researchers should also be wary of such an approach. In other words, further research is needed to tease apart these different response processes. Second, these findings may be applicable to computer adaptive testing in order to weed out early, unstable item responses. Once again, however, the current application would, not tenable for an adaptive investigation because items are grouped into only three blocks. Rather, a study in which each individual position is taken into account in the model might be more appropriate.

APPENDIX A: SAS Simulation Program

```
filename outdat 'C:\Heather\500_15_Le_Lf_2.DAT';
data blockit;
        do subject=1 to 500;
                do i=1 to 15;
                x=ranuni(555556);
                output;
                end;
        end;
run;
proc sort;
by subject x;
run;

data blockit;
set blockit;
by subject x;
if first.subject then sp=0;
sp+1;
if sp le 5 then do;
        block=1;
        mult=1;
        add=0;
end;
else if sp le 10 then do;
        block=2;
        mult=1.7;
        add=.15;
end;
```

```
else if sp le 15 then do;
        block=3;
        mult=1.8;
        add=.20;
end;
run;

proc sort;
by subject i;
proc print;
run;

proc transpose data=blockit out=blocktran1 prefix=block;
var block;
by subject;
id i;
run;
proc print;
run;

proc transpose data=blockit out=blocktran2 prefix=MULT;
var MULT;
by subject;
id i;
run;
proc print;
run;

proc transpose data=blockit out=blocktran3 prefix=ADD;
```

```
var ADD;
by subject;
id i;
run;
proc print;
run;

DATA BLOCKTRAN;
MERGE BLOCKTRAN1 BLOCKTRAN2 BLOCKTRAN3;
BY SUBJECT;
theta=rannor(555555);
DROP _NAME_;
RUN;
PROC PRINT;
RUN;

data ratsim;
set Blocktran;
by subject;
retain theta;
ncat=4; m=3;
array delta (i) d1-d15;
array alpha (i) a1-a15;
array response (i) r1-r15;
ARRAY MULT (I) MULT1-MULT15;
ARRAY ADD (I) ADD1-ADD15;

do i=1 to 15;
```

d1=-.10;

d2=-.56;

d3=.69;

d4=-.14;

d5=-1.17;

d6=.01;

d7=-1.14;

d8=.19;

d9=.08;

d10=-1.27;

d11=-.16;

d12=-.09;

d13=.58;

d14=-.22;

d15=.30;

a1=1.84;

a2=1.84;

a3=1.69;

a4=1.62;

a5=1.59;

a6=2.60;

a7=2.13;

a8=2.02;

a9=1.97;

a10=1.90;

a11=1.47;

a12=1.44;

```
a13=1.42;
a14=1.32;
a15=1.29;

    IF i=1 THEN DO;
        sb1=delta-1.12;
        sb2=delta+0;
        sb3=delta+.55;
    END;
        IF i=2 THEN DO;
        sb1=delta-.79;
        sb2=delta+0;
        sb3=delta+.04;
    END;
    IF i=3 THEN DO;
        sb1=delta-.49;
        sb2=delta+0;
        sb3=delta+.36;
    END;
        IF i=4 THEN DO;
        sb1=delta-1.29;
        sb2=delta+0;
        sb3=delta+.71;
    END;
    IF i=5 THEN DO;
        sb1=delta-1.24;
        sb2=delta+0;
        sb3=delta+1.13;
    END;
```

```
IF i=6 THEN DO;
    sb1=delta-1.96;
    sb2=delta+0;
    sb3=delta+.39;
END;
IF i=7 THEN DO;
    sb1=delta-1.32;
    sb2=delta+0;
    sb3=delta+1.12;
END;
    IF i=8 THEN DO;
    sb1=delta-.85;
    sb2=delta+0;
    sb3=delta+1.05;
END;
IF i=9 THEN DO;
    sb1=delta-1.02;
    sb2=delta+0;
    sb3=delta+.94;
END;
    IF i=10 THEN DO;
    sb1=delta-1.54;
    sb2=delta+0;
    sb3=delta+1.19;
END;
IF i=11 THEN DO;
    sb1=delta-1.56;
    sb2=delta+0;
    sb3=delta+1.03;
```

```
END;
IF i=12 THEN DO;
      sb1=delta-1.41;
      sb2=delta+0;
      sb3=delta+1.06;
END;
      IF i=13 THEN DO;
      sb1=delta-.15;
      sb2=delta+0;
      sb3=delta+.76;
END;
IF i=14 THEN DO;
      sb1=delta-1.16;
      sb2=delta+0;
      sb3=delta+.54;
END;
      IF i=15 THEN DO;
      sb1=delta-.57;
      sb2=delta+0;
      sb3=delta+1.62;
END;
        END;
        run;

probc0=1/(exp(x*alpha*mult*(theta-sb1-add)));
probc1=exp(x*alpha*mult*(theta-sb1-add));
probc2=exp(x*alpha*mult*(theta-sb2-add));
probc3=exp(x*alpha*mult*(theta-sb3-add));
end;
```

```
    den=sum(of probc1-probc3);

check=0;
            prob1=probc1/den;
            prob2=probc2/den;
            prob3=probc3/den;
    check=check+prob1;
    check=check+prob2;
    check=check+prob3

  if abs(check-1) gt .01 then put "PROBLEM WITH ITEM" I=;
  cp0=probc0;
  cp1=cp0+prob1;
  cp2=cp1+prob2;
  cp3=cp2+prob3;
  rr=ranuni(777776);
  if rr le cp0 then response=1;
  else if rr le cp1 then response=2;
  else if rr le cp2 then response=3;
  else if rr le cp3 then response=4;
end;
run;
proc print; run;
/**completed, simulated data file based on above commands**/
data _null_;
set;
file outdat noprint notitles;
put (r1-r15 block1-block15) (15*2. 15*2.);
run;
```

```
proc freq;
tables r1-r15 block1-block15;
run;
```

APPENDIX B: WinBUGS Parameter Estimation Program

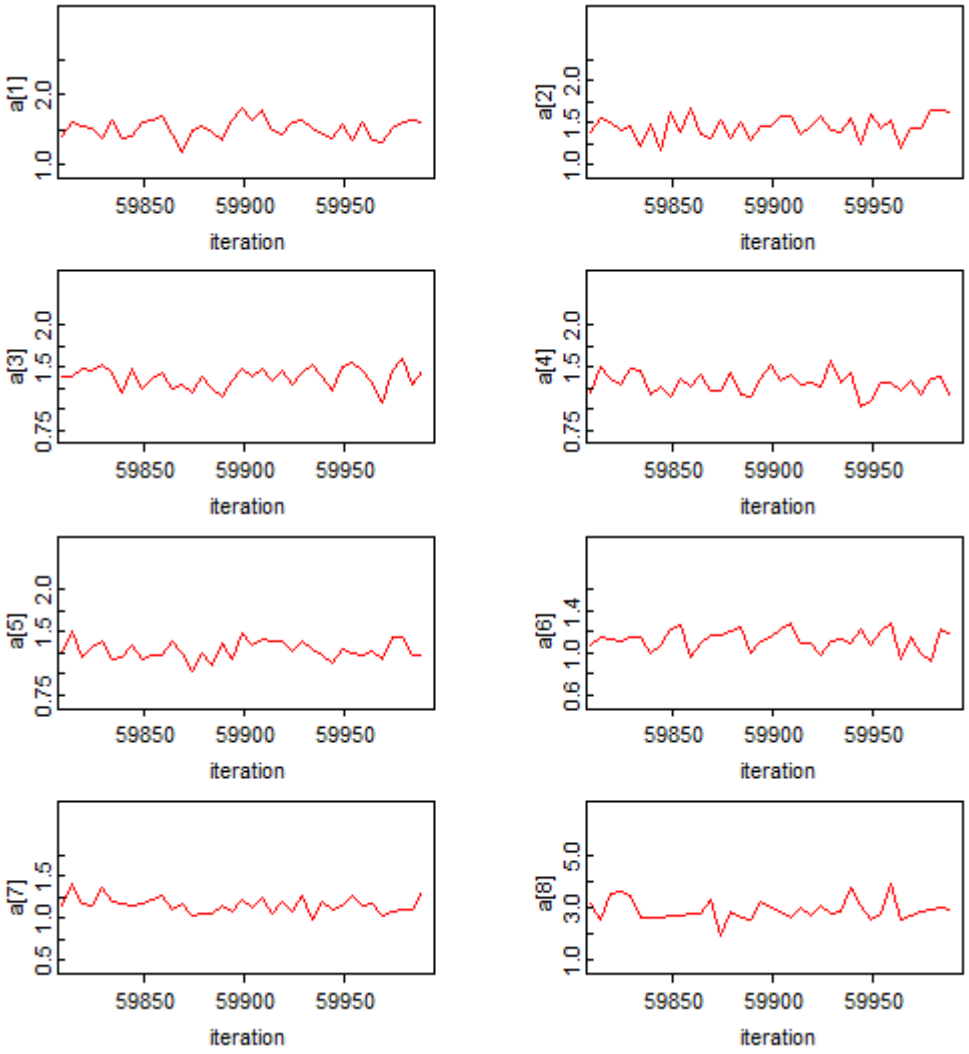
```
model GPCMF

{

#GPCFM 4 cat 20 items;

mult[1]<-1;

mult[2]~ dlnorm(0, 4);

mult[3]~ dlnorm(0, 4);

add[1]<-0;

add[2]~dnorm(0, .25);

add[3]~dnorm(0, .25);

for (k in 1:I)      {

          sb1[k] ~ dnorm(0, .25);

          sb2[k] ~ dnorm(0, .25);

          sb3[k] ~ dnorm(0, .25);

          a[k] ~ dlnorm(0, 4);

          }

for (j in 1:N)      {

for (k in 1:I)      {

pm[j,k,1]<- 1/(1+exp(a[k]*mult[s[j,k]]*((theta[j] - add[s[j,k]]) - sb1[k]))+
                    exp(a[k]*mult[s[j,k]]*(2*(theta[j] - add[s[j,k]]) - sb1[k] - sb2[k]))+
                              exp(a[k]*mult[s[j,k]]*(3*(theta[j] - add[s[j,k]]) - sb1[k] - sb2[k] -
sb3[k])));

pm[j,k,2]<- (exp(a[k]*mult[s[j,k]]*((theta[j] - add[s[j,k]]) - sb1[k])))/
(1+exp(a[k]*mult[s[j,k]]*((theta[j] - add[s[j,k]]) - sb1[k]))+
          exp(a[k]*mult[s[j,k]]*(2*(theta[j] - add[s[j,k]]) - sb1[k] - sb2[k]))+
                    exp(a[k]*mult[s[j,k]]*(3*(theta[j] - add[s[j,k]]) - sb1[k] - sb2[k] - sb3[k])));


pm[j,k,3]<- (exp(a[k]*mult[s[j,k]]*(2*(theta[j] - add[s[j,k]]) - sb1[k] - sb2[k])))/
(1+exp(a[k]*mult[s[j,k]]*((theta[j] - add[s[j,k]]) - sb1[k]))+
          exp(a[k]*mult[s[j,k]]*(2*(theta[j] - add[s[j,k]]) - sb1[k] - sb2[k]))+
                    exp(a[k]*mult[s[j,k]]*(3*(theta[j] - add[s[j,k]]) - sb1[k] - sb2[k] - sb3[k])));
```

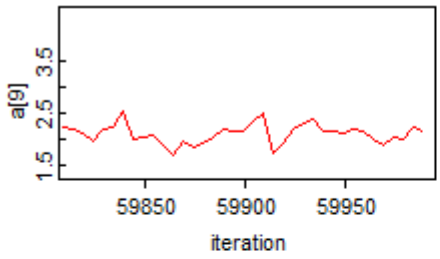
```
pm[j,k,4]<- (exp(a[k]*mult[s[j,k]]*(3*(theta[j] - add[s[j,k]]) - sb1[k] - sb2[k] - sb3[k])))/
(1+exp(a[k]*mult[s[j,k]]*((theta[j] - add[s[j,k]]) - sb1[k]))+
        exp(a[k]*mult[s[j,k]]*(2*(theta[j] - add[s[j,k]]) - sb1[k] - sb2[k]))+
                exp(a[k]*mult[s[j,k]]*(3*(theta[j] - add[s[j,k]]) - sb1[k] - sb2[k] - sb3[k])));

r[j,k] ~ dcat(pm[j,k, ]);

        }

        theta[j] ~ dnorm(0, 1);

        }

        }

list(N=2000,I=15)

r[,1] r[,2] r[,3] r[,4] r[,5] r[,6] r[,7] r[,8] r[,9] r[,10] r[,11] r[,12] r[,13] r[,14] r[,15]

s[,1] s[,2] s[,3] s[,4] s[,5] s[,6] s[,7] s[,8] s[,9] s[,10] s[,11] s[,12] s[,13] s[,14] s[,15]

 4 4 4 4 4 4 4 4 2 3 3 2 1 4 1
```

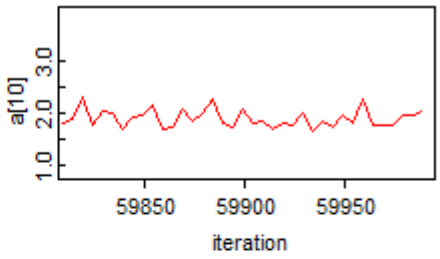
APPENDIX C: Pilot Simulation Description

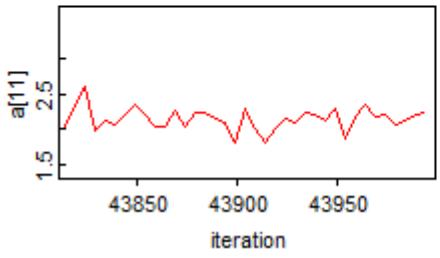
Data for two four two cells, in which model parameters had simplistic real values, were submitted to five replications each. Specifically, both cells consisted of an N of 2000, I= 30, with the first cell being composed of small e_k and f_k values and the second cell having large e_k and f_k values. For the first cell, the average e_1 and e_2 were 1.049 and 1.099, respectively, across replications, while the average e_1 and e_2 for the second cell were 1.789 and 1.899, respectively, across replications. With respect to f_1 and f_2, on average (across replications), the estimated parameters were f_1 =.049 and f_2 =.109, respectively, for cell 1, and f_1 =.156 and f_2 =.201, respectively, for cell 2. With the exception of θ, which was randomly sampled from a normal distribution, ~N(0,1), for every replication, the remaining parameters in the model were kept constant across all cells and replications. Based on the winBUGS analysis and initial examinations of model convergence (e.g., trace plots) and parameter recovery for all ten replication outputs, the model appears to be viable. See Figures 2a through 2d for a randomly sampled replication (cell 1, replication 5) that demonstrates the relationship between true and estimated values for each parameter. Other replications yielded almost identical results to those shown in Figure 2. Therefore, continuation of this project is deemed appropriate.

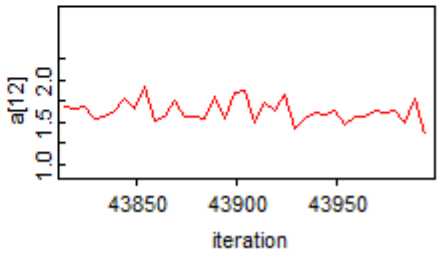
APPENDIX D: Trace Plots

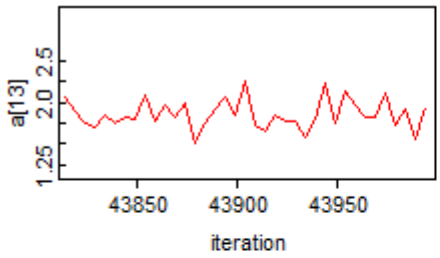


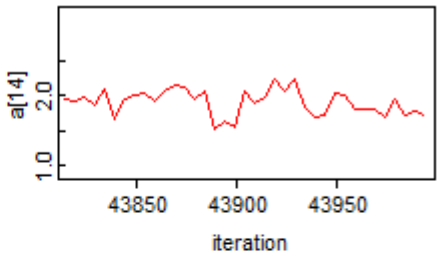

a[9]
1.5
2.5
3.5
59850
59900
59950
iteration

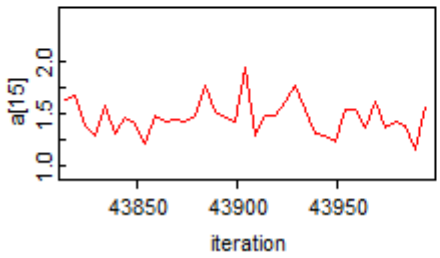

a[10]
1.0
2.0
3.0
59850
59900
59950
iteration

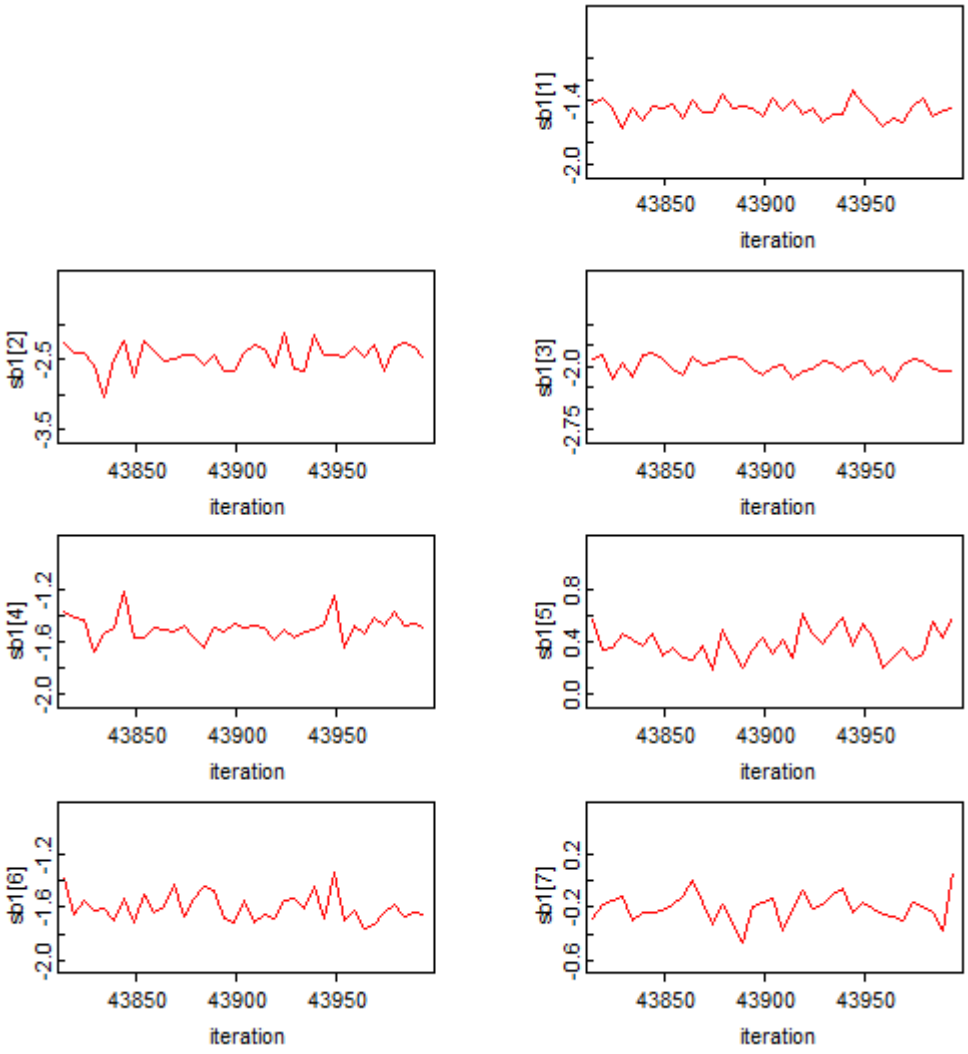

a[11]
1.5
2.5
43850
43900
43950
iteration

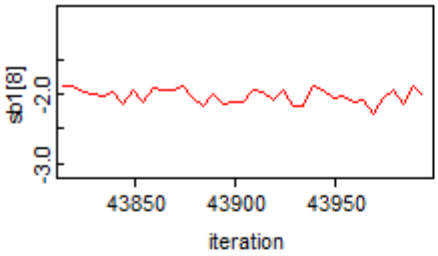

a[12]
1.0
1.5
2.0
43850
43900
43950
iteration

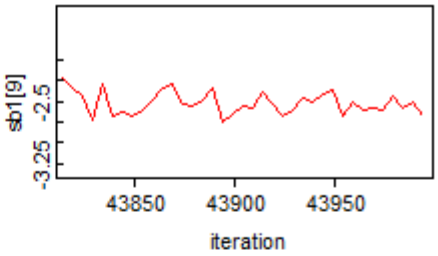

a[13]
1.25
2.0
2.5
43850
43900
43950
iteration

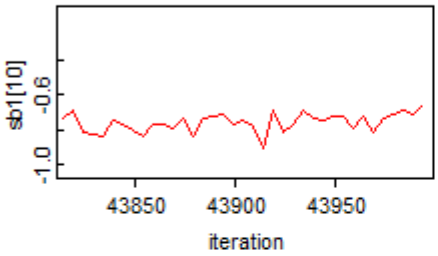

a[14]
1.0
2.0
43850
43900
43950
iteration

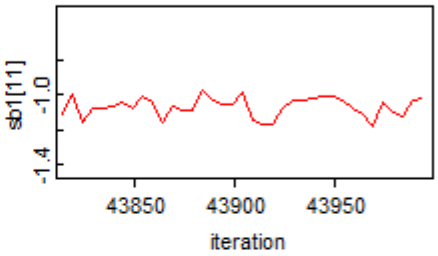

a[15]
1.0
1.5
2.0
43850
43900
43950
iteration

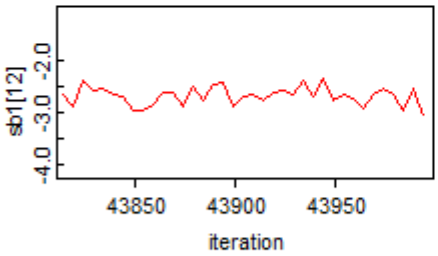

sb1[1]
-2.0
-1.4
43850
43900
43950
iteration
sb1[2]
-3.5
-2.5
43850
43900
43950
iteration
sb1[3]
-2.75
-2.0
43850
43900
43950
iteration
sb1[4]
-2.0
-1.6
-1.2
43850
43900
43950
iteration
sb1[5]
0.0
0.4
0.8
43850
43900
43950
iteration
sb1[6]
-2.0
-1.6
-1.2
43850
43900
43950
iteration
sb1[7]
-0.6
-0.2
0.2
43850
43900
43950
iteration

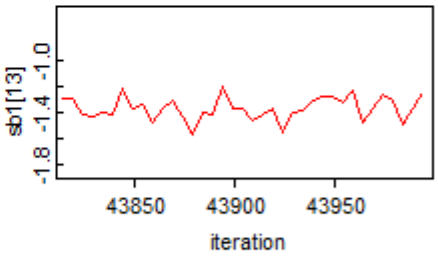

sb1[8]
-3.0
-2.0
43850
43900
43950
iteration

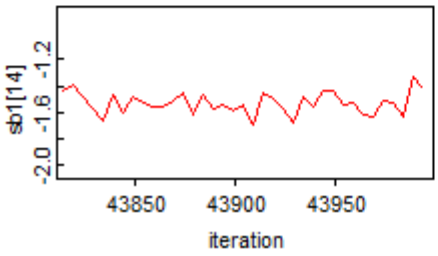

sb1[9]
-3.25
-2.5
43850
43900
43950
iteration

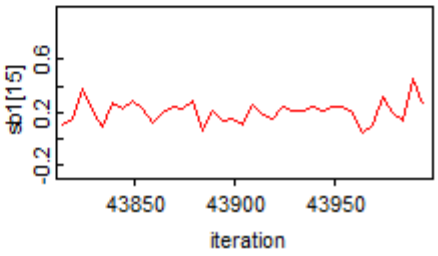

sb1[10]
-1.0
-0.6
43850
43900
43950
iteration

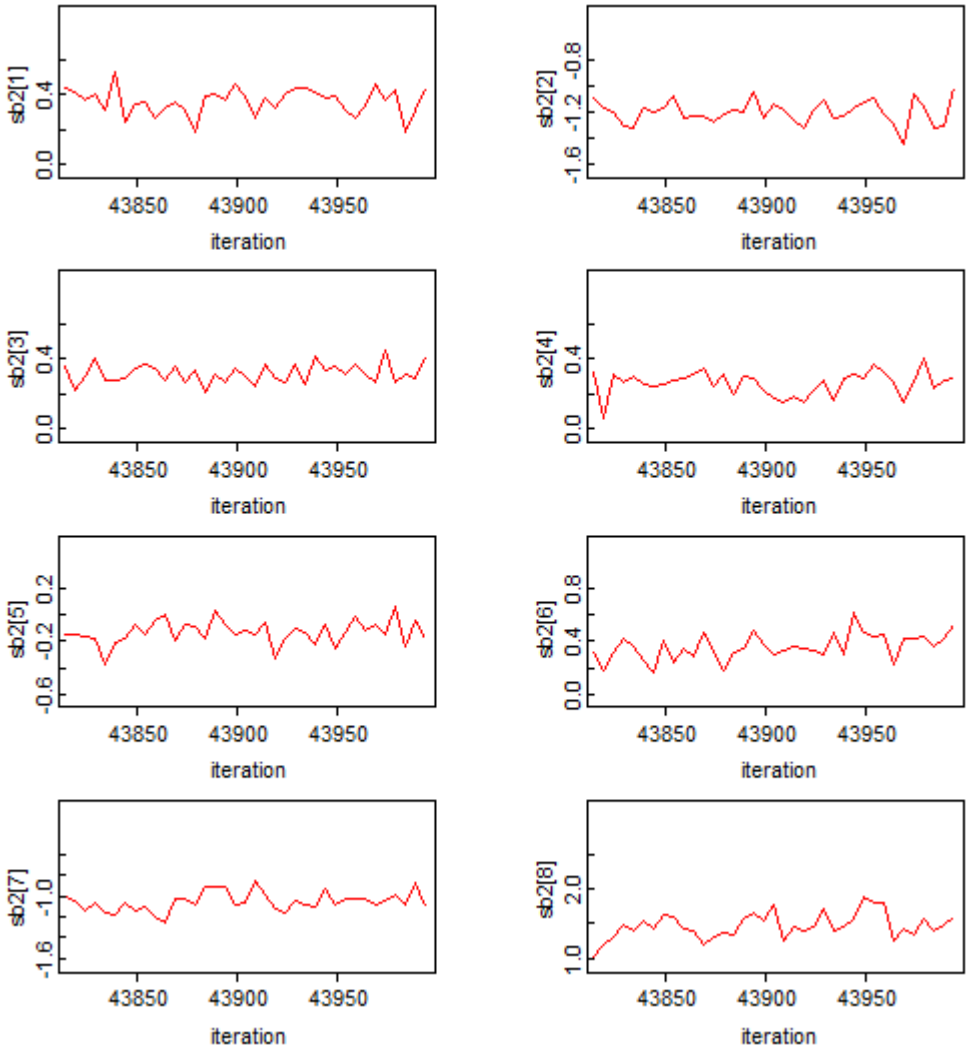

sb1[11]
-1.4
-1.0
43850
43900
43950
iteration

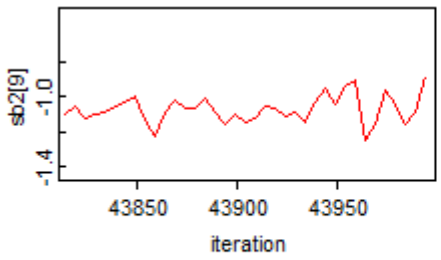

sb1[12]
-4.0
-3.0
-2.0
43850
43900
43950
iteration

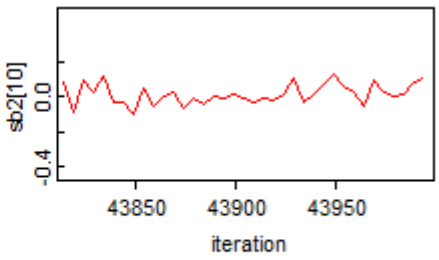

sb1[13]
-1.8
-1.4
-1.0
43850
43900
43950
iteration

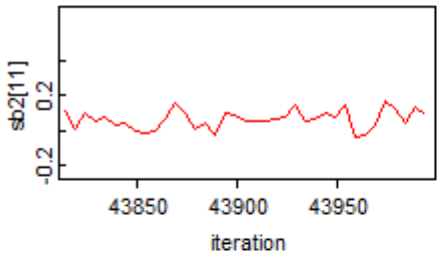

sb1[14]
-2.0
-1.6
-1.2
43850
43900
43950
iteration

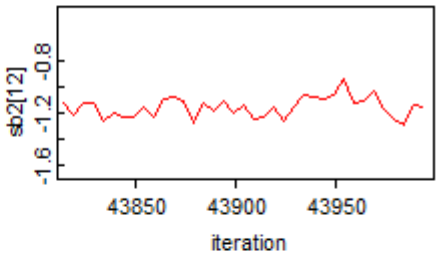

sb1[15]
-0.2
0.2
0.6
43850
43900
43950
iteration

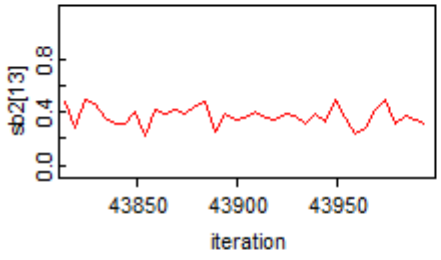

sb2[1]
0.0
0.4
43850
43900
43950
iteration
sb2[2]
-1.6
-1.2
-0.8
43850
43900
43950
iteration
sb2[3]
0.0
0.4
43850
43900
43950
iteration
sb2[4]
0.0
0.4
43850
43900
43950
iteration
sb2[5]
-0.6
-0.2
0.2
43850
43900
43950
iteration
sb2[6]
0.0
0.4
0.8
43850
43900
43950
iteration
sb2[7]
-1.6
-1.0
43850
43900
43950
iteration
sb2[8]
1.0
2.0
43850
43900
43950
iteration

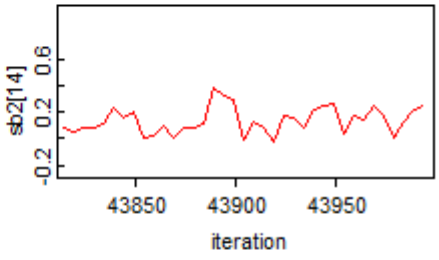

sb2[9]
-1.4
-1.0
43850
43900
43950
iteration

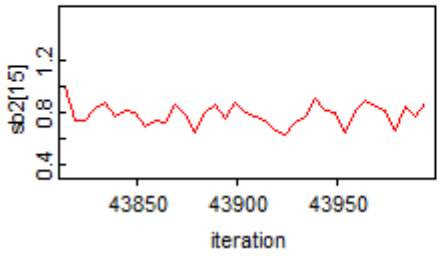

sb2[10]
-0.4
0.0
43850
43900
43950
iteration

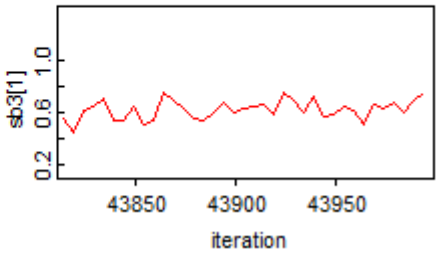

sb2[11]
-0.2
0.2
43850
43900
43950
iteration

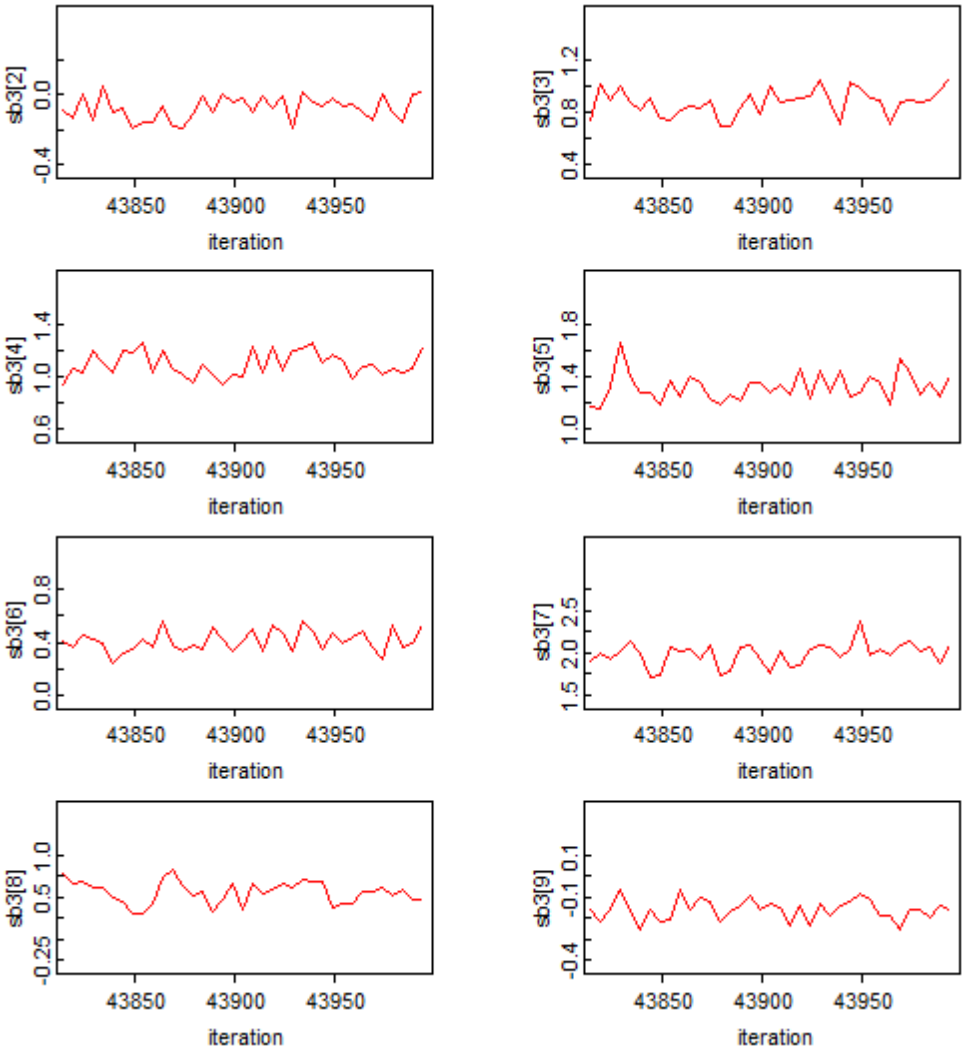

sb2[12]
-1.6
-1.2
-0.8
43850
43900
43950
iteration

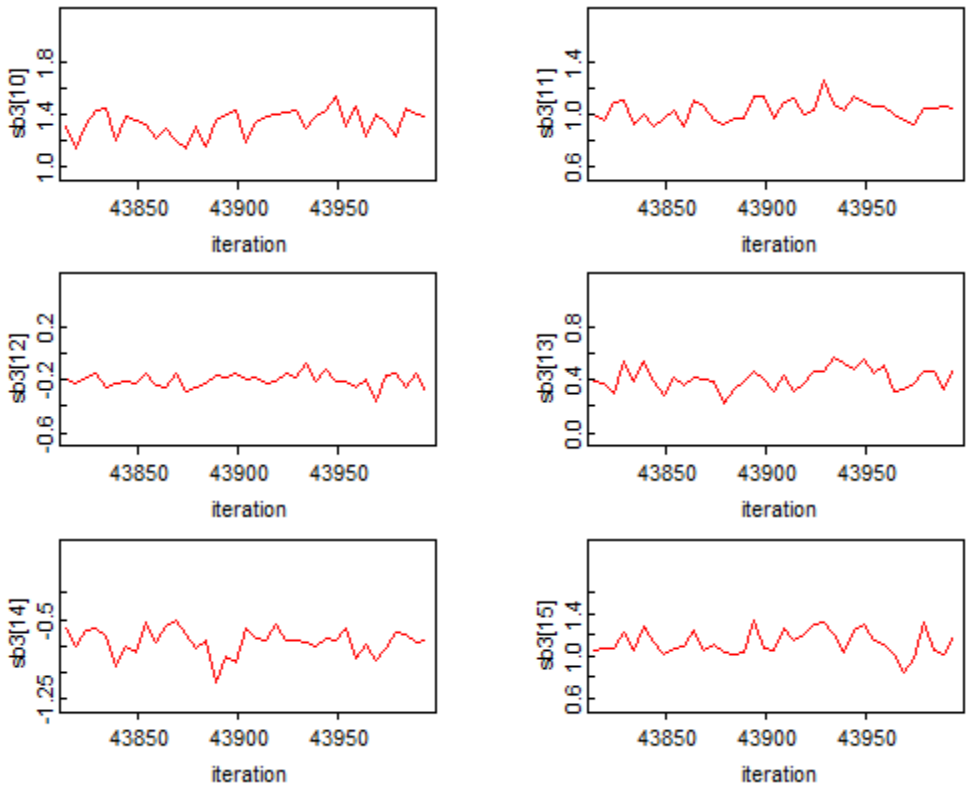

sb2[13]
0.0
0.4
0.8
43850
43900
43950
iteration


sb2[14]
-0.2
0.2
0.6
43850
43900
43950
iteration

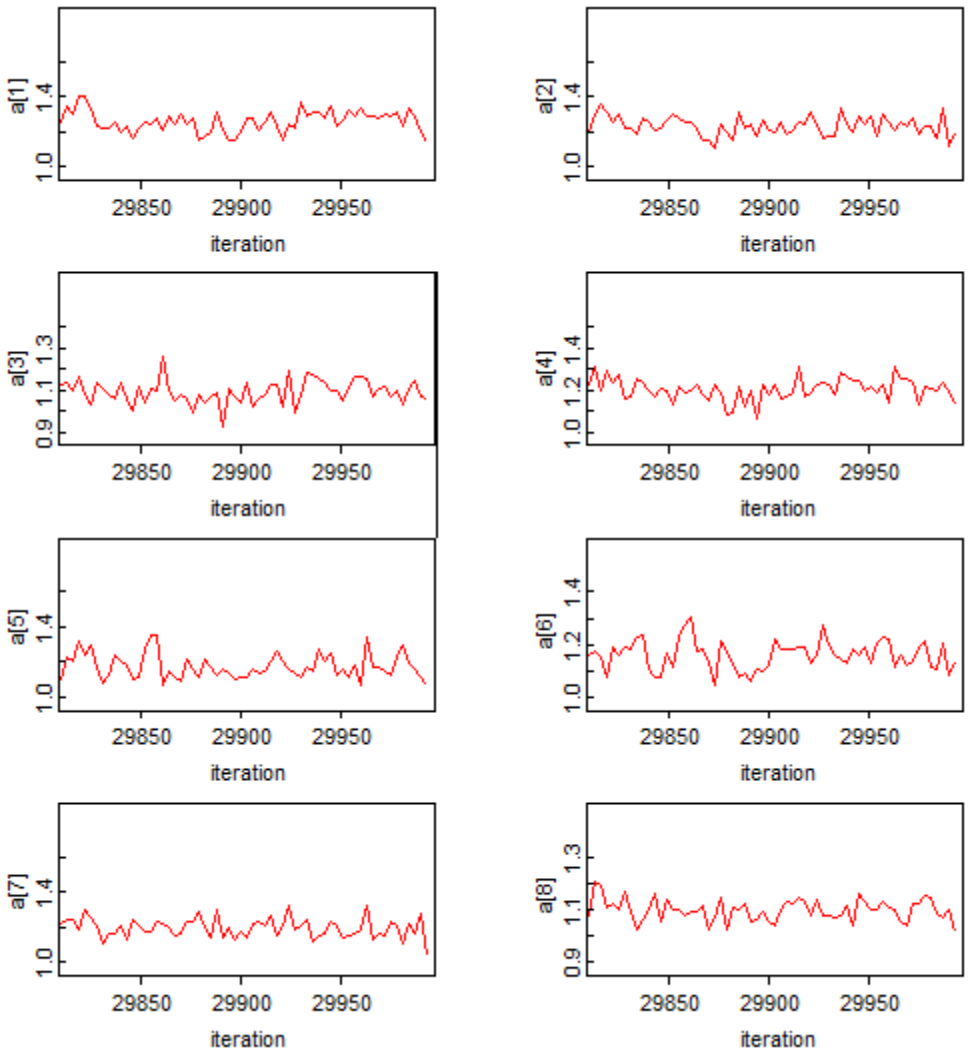

sb2[15]
0.4
0.8
1.2
43850
43900
43950
iteration

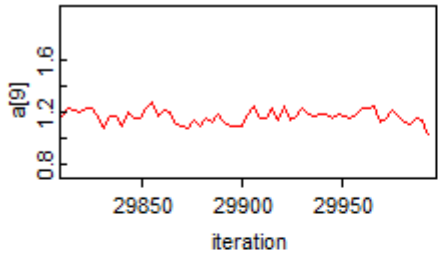

sb3[1]
0.2
0.6
1.0
43850
43900
43950
iteration

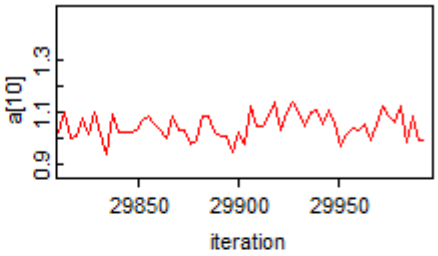

sb3[2]
-0.4
0.0
sb3[3]
0.4
0.8
1.2
sb3[4]
0.6
1.0
1.4
sb3[5]
1.0
1.4
1.8
sb3[6]
0.0
0.4
0.8
sb3[7]
1.5
2.0
2.5
sb3[8]
-0.25
0.5
1.0
sb3[9]
-0.4
-0.1
0.1
43850
43900
43950
iteration

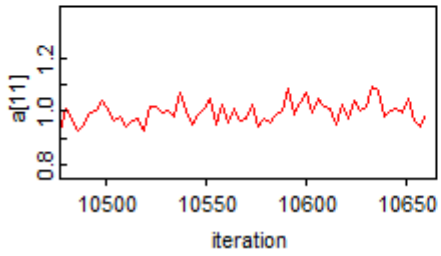


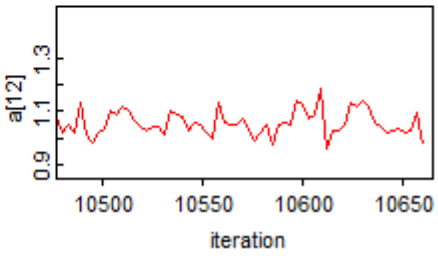
N=500, I=15, Large e, Large f

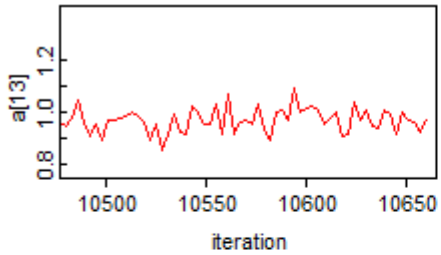

a[1]
1.0
1.4
29850
29900
29950
iteration
a[2]
1.0
1.4
29850
29900
29950
iteration
a[3]
0.9
1.1
1.3
29850
29900
29950
iteration
a[4]
1.0
1.2
1.4
29850
29900
29950
iteration
a[5]
1.0
1.4
29850
29900
29950
iteration
a[6]
1.0
1.2
1.4
29850
29900
29950
iteration
a[7]
1.0
1.4
29850
29900
29950
iteration
a[8]
0.9
1.1
1.3
29850
29900
29950
iteration

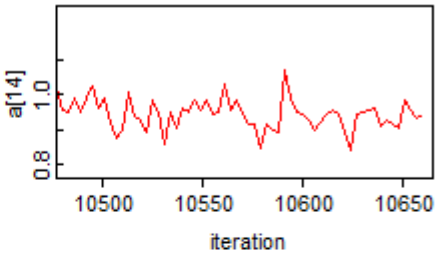

a[9]
0.8
1.2
1.6
29850
29900
29950
iteration

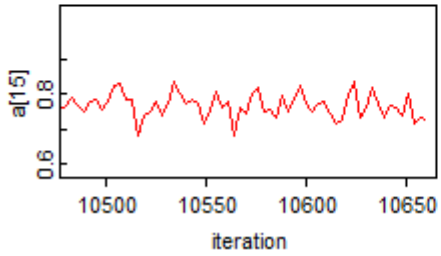

a[10]
0.9
1.1
1.3
29850
29900
29950
iteration

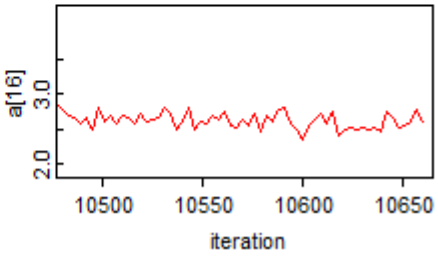

a[11]
0.8
1.0
1.2
10500
10550
10600
10650
iteration

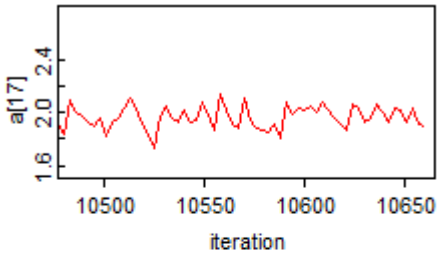

a[12]
0.9
1.1
1.3
10500
10550
10600
10650
iteration

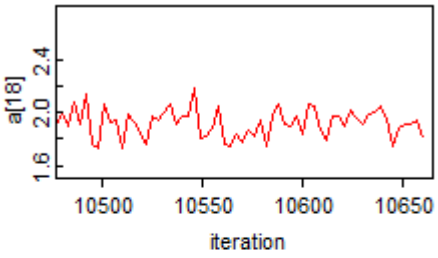

a[13]
0.8
1.0
1.2
10500
10550
10600
10650
iteration

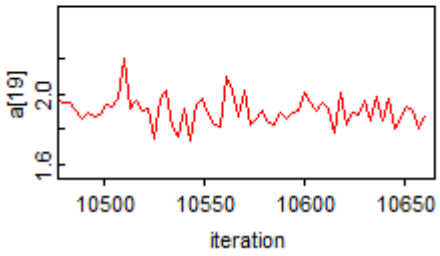

a[14]
0.8
1.0
10500
10550
10600
10650
iteration

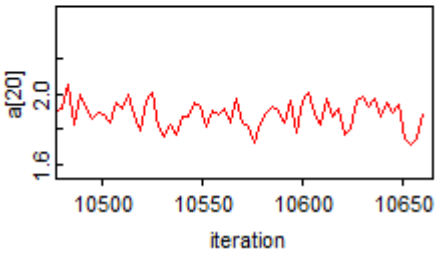

a[15]
0.6
0.8
10500
10550
10600
10650
iteration

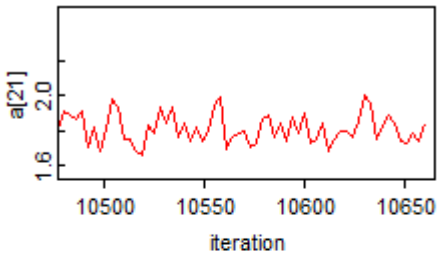

a[16]
2.0
3.0
10500
10550
10600
10650
iteration

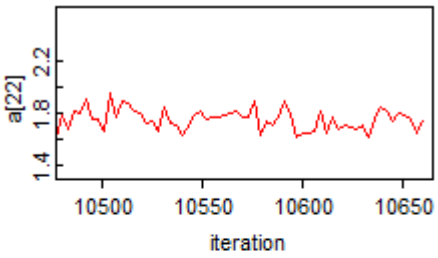

a[17]
1.6
2.0
2.4
10500
10550
10600
10650
iteration

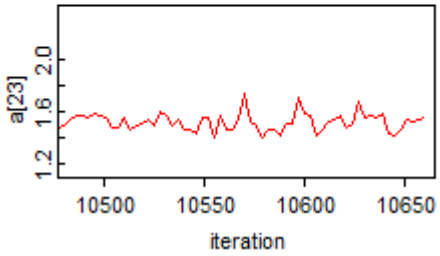

a[18]
1.6
2.0
2.4
10500
10550
10600
10650
iteration

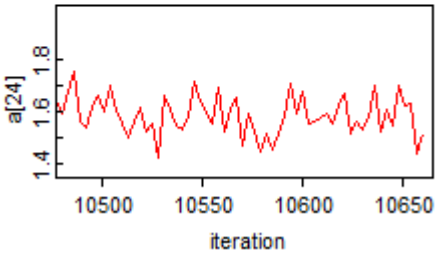

a[19]
1.6
2.0
10500
10550
10600
10650
iteration

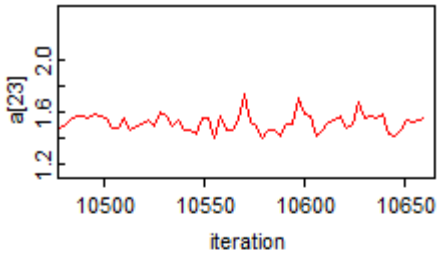

a[20]
1.6
2.0
10500
10550
10600
10650
iteration

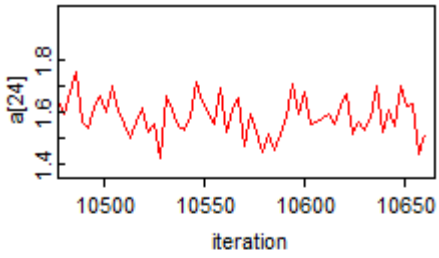

a[21]
1.6
2.0
10500
10550
10600
10650
iteration

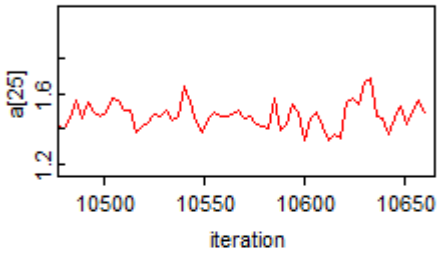

a[22]
1.4
1.8
2.2
10500
10550
10600
10650
iteration

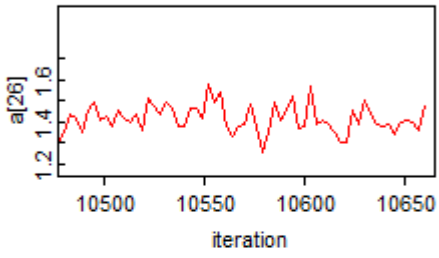

a[23]
1.2
1.6
2.0
10500
10550
10600
10650
iteration

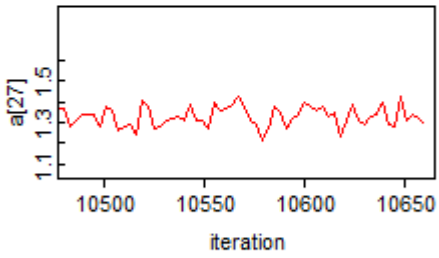

a[24]
1.4
1.6
1.8
10500
10550
10600
10650
iteration

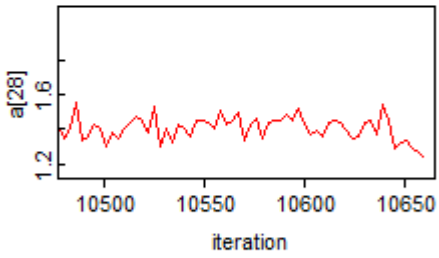

a[23]
1.2 1.6 2.0
10500 10550 10600 10650
iteration

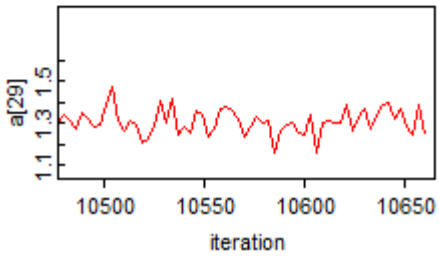

a[24]
1.4 1.6 1.8
10500 10550 10600 10650
iteration

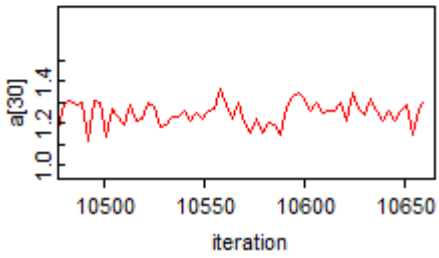

a[25]
1.2 1.6
10500 10550 10600 10650
iteration

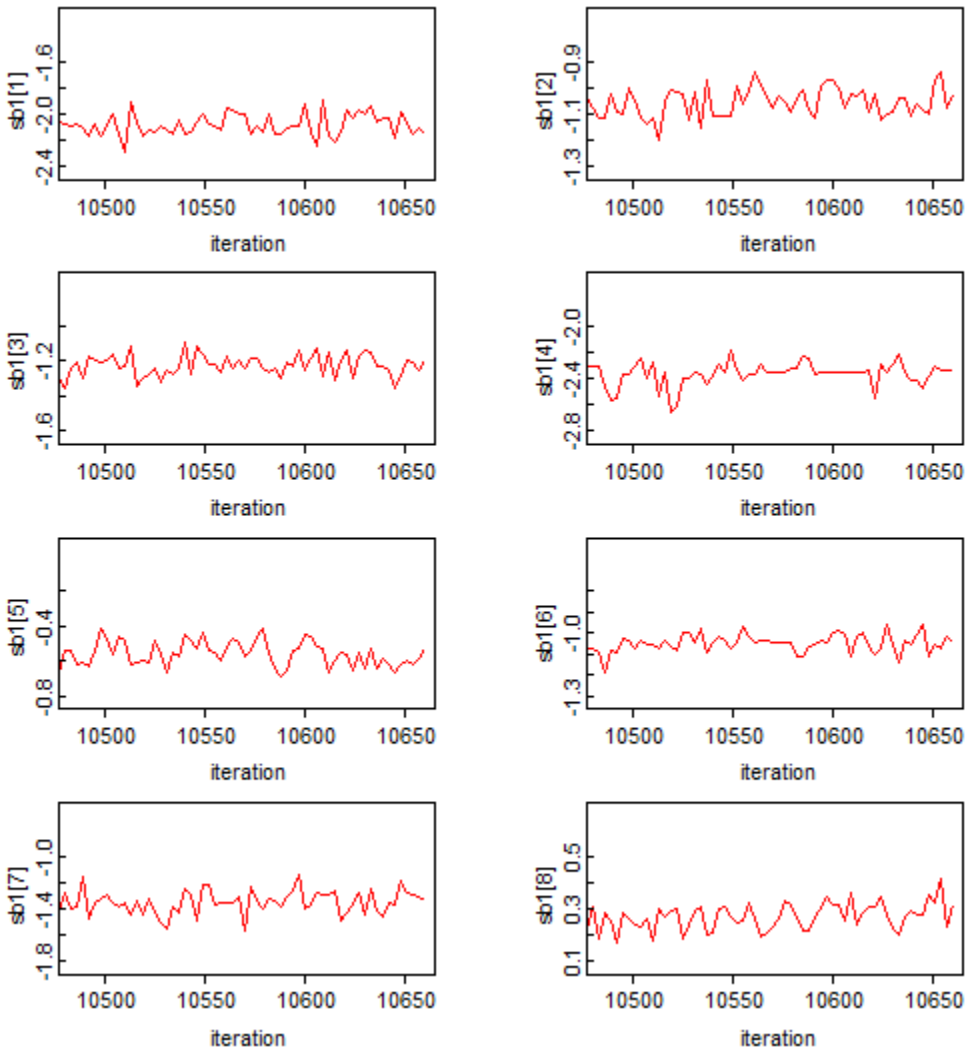

a[26]
1.2 1.4 1.6
10500 10550 10600 10650
iteration

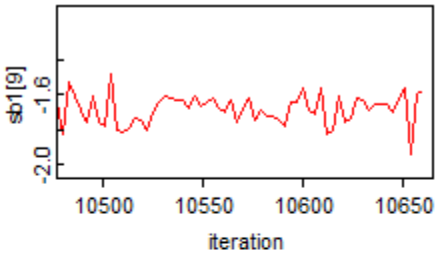

a[27]
1.1 1.3 1.5
10500 10550 10600 10650
iteration

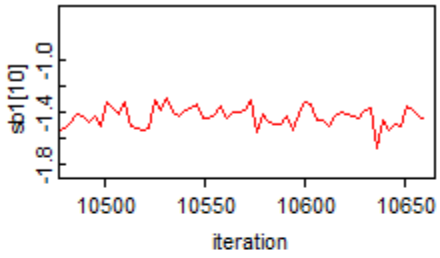

a[28]
1.2 1.6
10500 10550 10600 10650
iteration

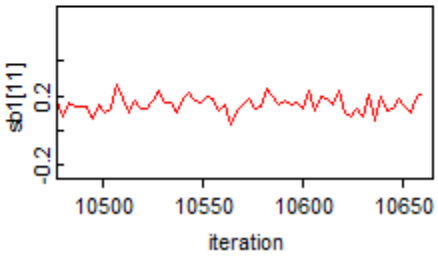

a[29]
1.1 1.3 1.5
10500 10550 10600 10650
iteration

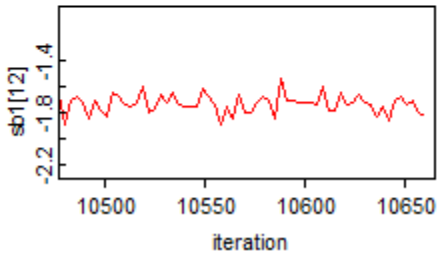

a[30]
1.0 1.2 1.4
10500 10550 10600 10650
iteration

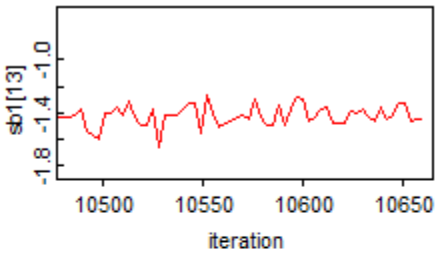

sb1[1]
-2.4 -2.0 -1.6
10500 10550 10600 10650
iteration
sb1[2]
-1.3 -1.1 -0.9
10500 10550 10600 10650
iteration
sb1[3]
-1.6 -1.2
10500 10550 10600 10650
iteration
sb1[4]
-2.8 -2.4 -2.0
10500 10550 10600 10650
iteration
sb1[5]
-0.8 -0.4
10500 10550 10600 10650
iteration
sb1[6]
-1.3 -1.0
10500 10550 10600 10650
iteration
sb1[7]
-1.8 -1.4 -1.0
10500 10550 10600 10650
iteration
sb1[8]
0.1 0.3 0.5
10500 10550 10600 10650
iteration

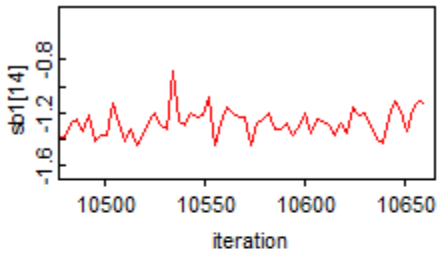

sb1[9]
-2.0
-1.6
10500
10550
10600
10650
iteration

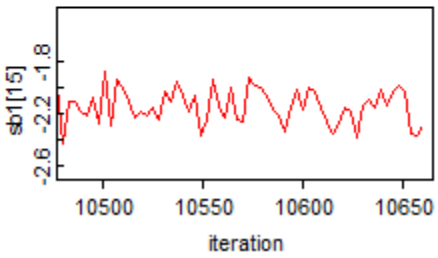

sb1[10]
-1.8
-1.4
-1.0
10500
10550
10600
10650
iteration

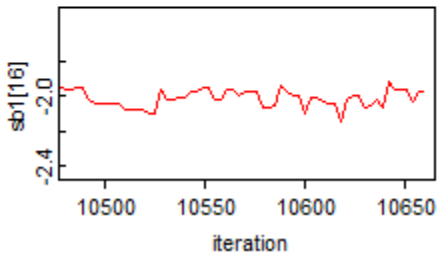

sb1[11]
-0.2
0.2
10500
10550
10600
10650
iteration

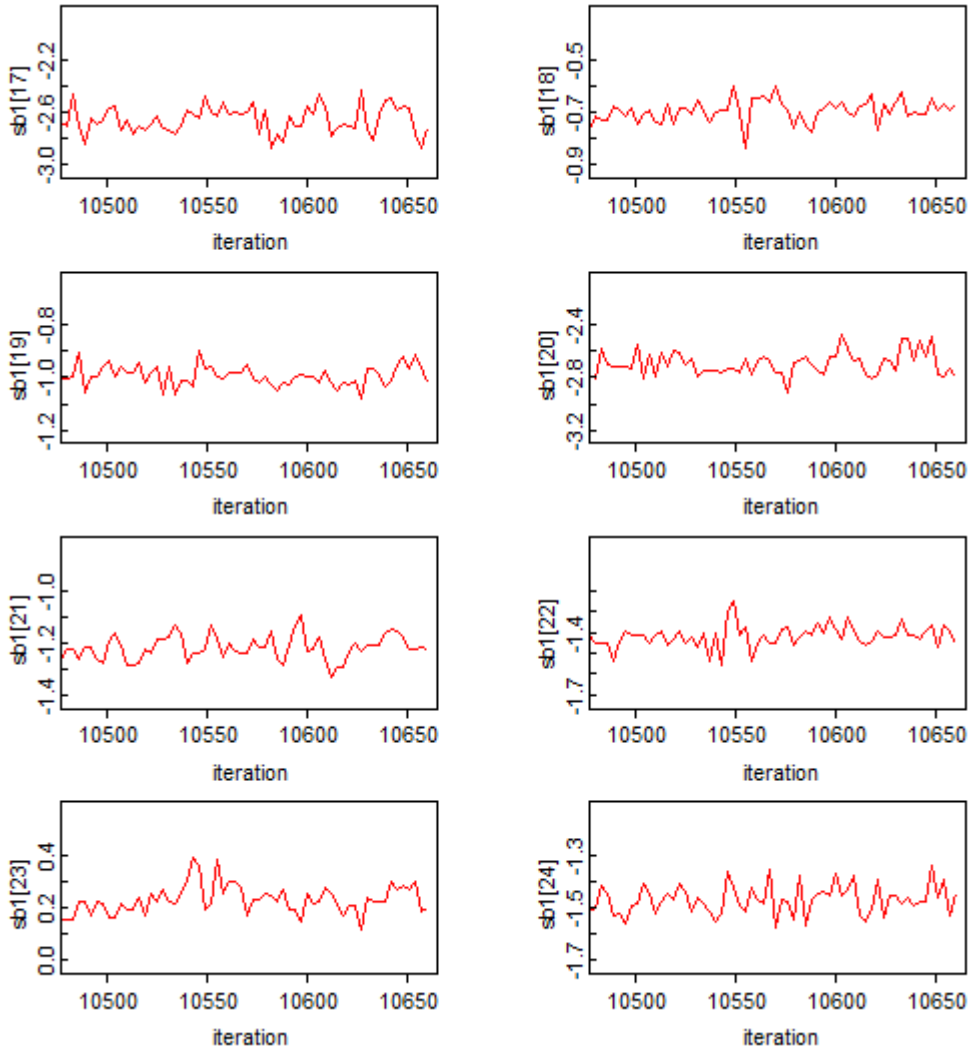

sb1[12]
-2.2
-1.8
-1.4
10500
10550
10600
10650
iteration

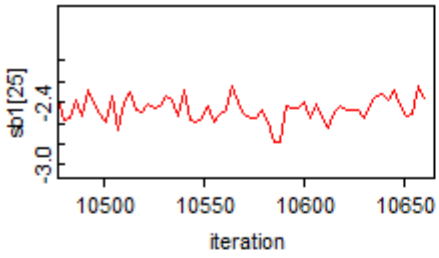

sb1[13]
-1.8
-1.4
-1.0
10500
10550
10600
10650
iteration

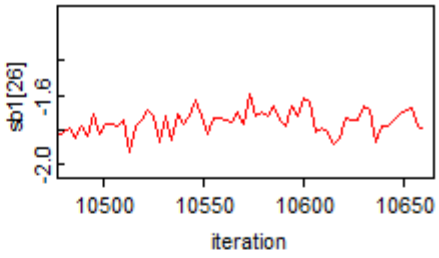

sb1[14]
-1.6
-1.2
-0.8
10500
10550
10600
10650
iteration

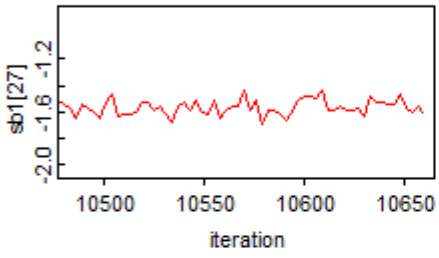

sb1[15]
-2.6
-2.2
-1.8
10500
10550
10600
10650
iteration

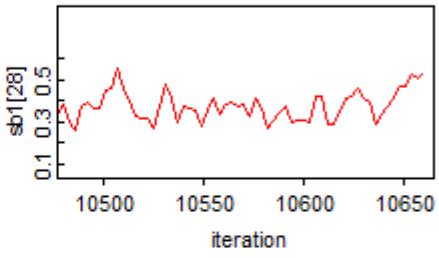

sb1[16]
-2.4
-2.0
10500
10550
10600
10650
iteration

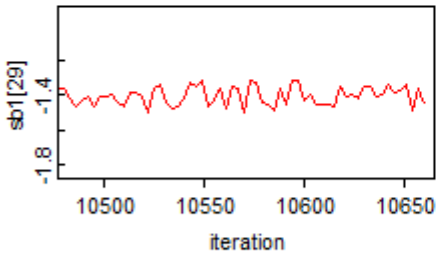

sb1[17]
-3.0 -2.6 -2.2
10500 10550 10600 10650
iteration
sb1[18]
-0.9 -0.7 -0.5
10500 10550 10600 10650
iteration
sb1[19]
-1.2 -1.0 -0.8
10500 10550 10600 10650
iteration
sb1[20]
-3.2 -2.8 -2.4
10500 10550 10600 10650
iteration
sb1[21]
-1.4 -1.2 -1.0
10500 10550 10600 10650
iteration
sb1[22]
-1.7 -1.4
10500 10550 10600 10650
iteration
sb1[23]
0.0 0.2 0.4
10500 10550 10600 10650
iteration
sb1[24]
-1.7 -1.5 -1.3
10500 10550 10600 10650
iteration

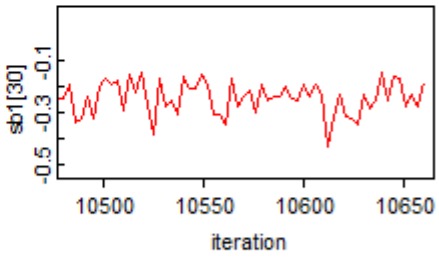

sb1[25]
-3.0
-2.4
10500
10550
10600
10650
iteration

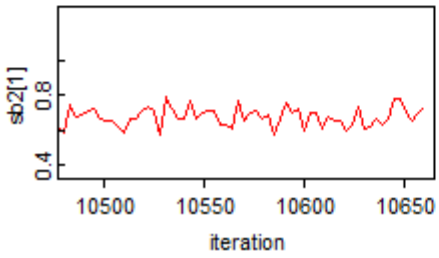

sb1[26]
-2.0
-1.6
10500
10550
10600
10650
iteration

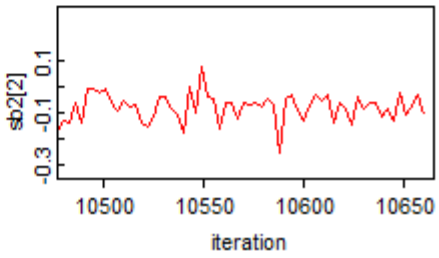

sb1[27]
-2.0
-1.6
-1.2
10500
10550
10600
10650
iteration

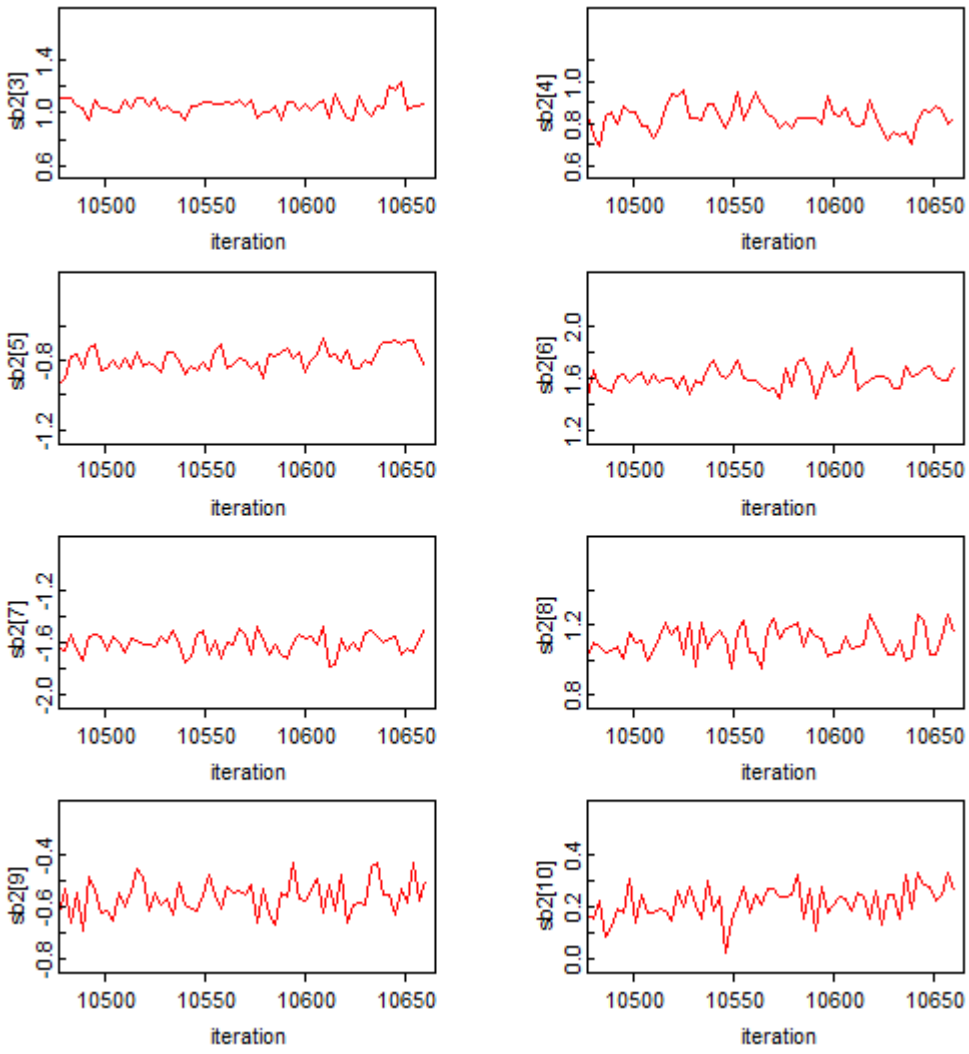

sb1[28]
0.1
0.3
0.5
10500
10550
10600
10650
iteration

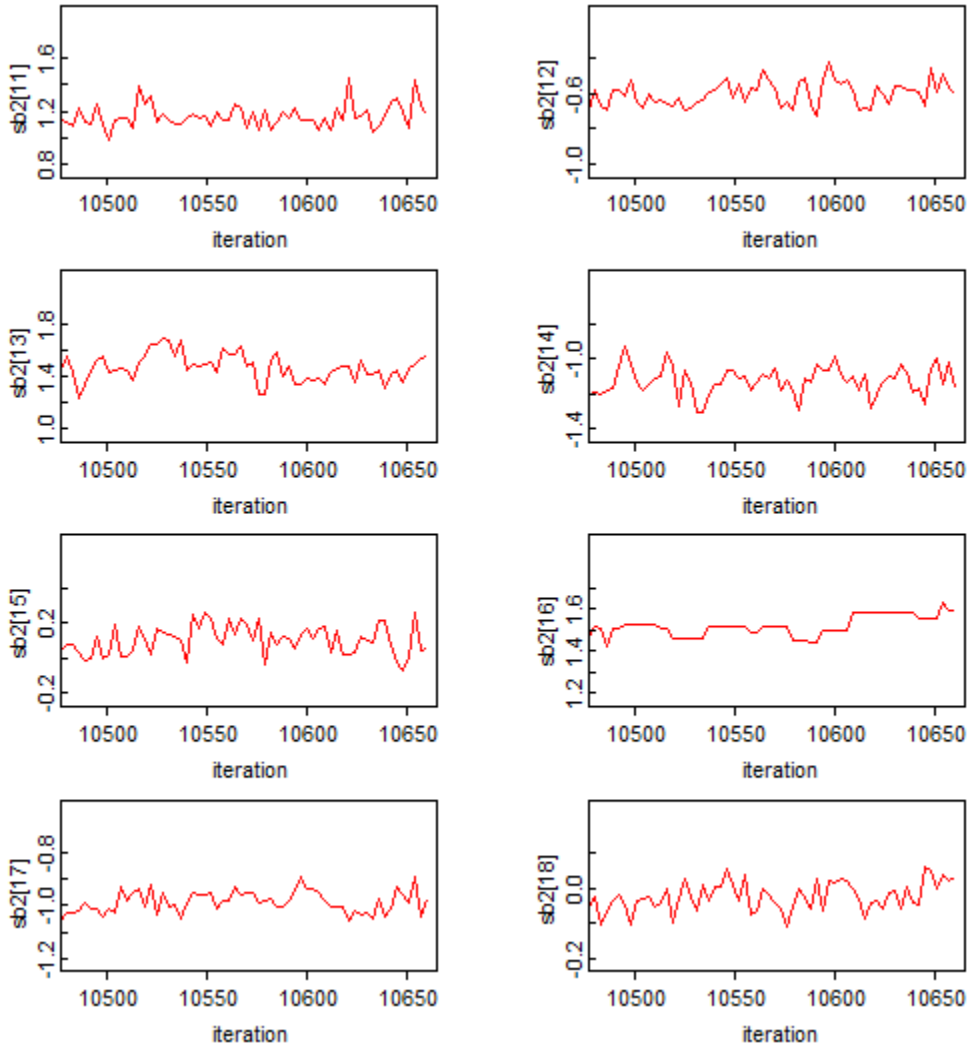

sb1[29]
-1.8
-1.4
10500
10550
10600
10650
iteration

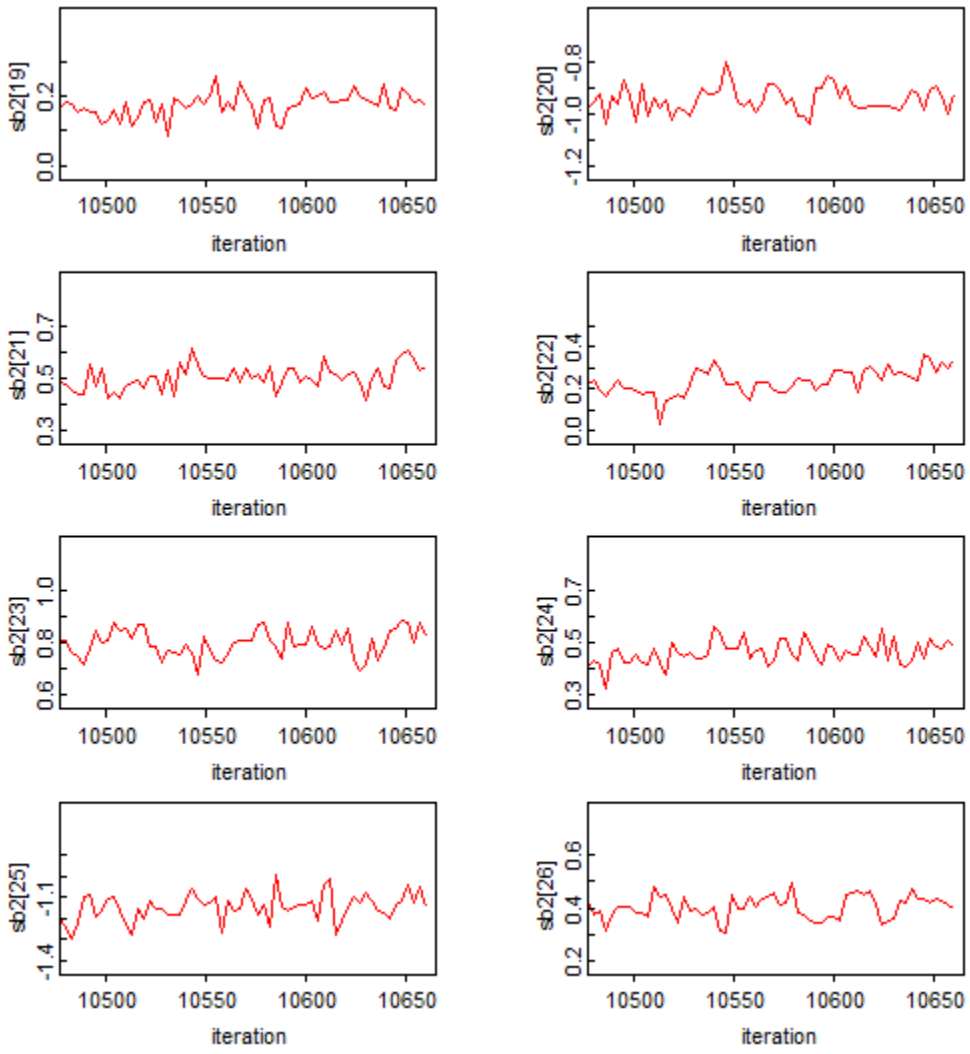

sb1[30]
-0.5
-0.3
-0.1
10500
10550
10600
10650
iteration


sb2[1]
0.4
0.8
10500
10550
10600
10650
iteration


sb2[2]
-0.3
-0.1
0.1
10500
10550
10600
10650
iteration


sb2[3]
0.6 1.0 1.4
10500 10550 10600 10650
iteration
sb2[4]
0.6 0.8 1.0
10500 10550 10600 10650
iteration
sb2[5]
-1.2 -0.8
10500 10550 10600 10650
iteration
sb2[6]
1.2 1.6 2.0
10500 10550 10600 10650
iteration
sb2[7]
-2.0 -1.6 -1.2
10500 10550 10600 10650
iteration
sb2[8]
0.8 1.2
10500 10550 10600 10650
iteration
sb2[9]
-0.8 -0.6 -0.4
10500 10550 10600 10650
iteration
sb2[10]
0.0 0.2 0.4
10500 10550 10600 10650
iteration


sb2[11]
0.8 1.2 1.6
sb2[12]
-1.0 -0.6
sb2[13]
1.0 1.4 1.8
sb2[14]
-1.4 -1.0
sb2[15]
-0.2 0.2
sb2[16]
1.2 1.4 1.6
sb2[17]
-1.2 -1.0 -0.8
sb2[18]
-0.2 0.0
10500 10550 10600 10650
iteration



N=2000, I=30, Small e, Small f

APPENDIX E: Item-Level MCMC Estimates and Errors by Parameter and Condition

Table E1a: i=15; α, small *e*, small *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	1.84	1.66	.06	.15	.19	1.79	.10	.11	.11	1.78	.01	.08	.06
2	1.84	1.76	.23	.17	.24	1.80	.11	.13	.11	1.79	.08	.08	.09
3	1.69	1.65	.17	.16	.18	1.67	.09	.11	.10	1.66	.09	.08	.09
4	1.62	1.47	.12	.13	.12	1.51	.05	.09	.12	1.54	.06	.07	.10
5	1.59	1.35	.05	.13	.25	1.44	.06	.10	.15	1.50	.06	.07	.10
6	2.60	2.51	.38	.30	.39	2.59	.16	.23	.16	2.60	.17	.15	.17
7	2.13	1.78	.13	.17	.37	1.93	.10	.13	.22	2.00	.12	.09	.17
8	2.02	1.78	.26	.16	.36	1.88	.06	.12	.14	1.94	.03	.09	.08
9	1.97	1.68	.16	.15	.33	1.84	.06	.11	.13	1.85	.05	.08	.13
10	1.90	1.72	.16	.16	.24	1.80	.13	.12	.16	1.87	.12	.09	.12
11	1.28	1.39	.10	.12	.15	1.45	.05	.09	.17	1.46	.03	.07	.18
12	1.27	1.28	.04	.11	.04	1.32	.07	.08	.08	1.39	.08	.06	.12
13	1.26	1.21	.14	.11	.16	1.33	.08	.09	.11	1.38	.06	.06	.14
14	1.26	1.18	.09	.11	.12	1.31	.04	.08	.10	1.28	.05	.06	.05
15	1.24	1.19	.14	.11	.15	1.24	.01	.08	.02	1.26	.03	.06	.03
total	1.66	1.57	.14	.15	.16	1.66	.07	.11	.07	1.69	.06	.08	.06

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E1b: i=15; α, small *e*, large *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	1.84	1.62	.11	.15	.13	1.76	.11	.11	.13	1.76	.06	.08	.10
2	1.84	1.69	.22	.17	.26	1.76	.13	.12	.15	1.76	.10	.08	.12
3	1.69	1.59	.11	.16	.14	1.67	.11	.12	.11	1.65	.08	.08	.08
4	1.62	1.45	.12	.13	.20	1.50	.05	.10	.13	1.53	.07	.07	.11
5	1.59	1.39	.07	.13	.21	1.47	.07	.10	.13	1.51	.05	.07	.09
6	2.60	2.51	.20	.30	.21	2.58	.15	.23	.15	2.60	.16	.16	.16
7	2.13	1.72	.13	.16	.43	1.91	.12	.13	.25	1.99	.10	.09	.17
8	2.02	1.82	.22	.17	.29	1.89	.10	.12	.16	1.96	.10	.09	.11
9	1.97	1.73	.13	.16	.27	1.88	.08	.12	.12	1.88	.03	.08	.09
10	1.90	1.75	.20	.16	.25	1.79	.16	.12	.19	1.86	.14	.09	.14
11	1.28	1.40	.07	.13	.13	1.45	.07	.09	.18	1.46	.05	.07	.18
12	1.27	1.29	.07	.12	.07	1.34	.08	.09	.10	1.39	.08	.06	.05
13	1.26	1.22	.13	.12	.13	1.32	.10	.09	.11	1.37	.05	.06	.12
14	1.26	1.25	.10	.11	.10	1.29	.04	.08	.05	1.26	.05	.06	.05
15	1.24	1.15	.07	.11	.11	1.24	.04	.08	.04	1.26	.04	.06	.04
total	1.66	1.57	.13	.15	.15	1.66	.09	.11	.09	1.68	.07	.08	.07

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E1c: i=15; α, large *e*, small *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	1.84	1.69	.15	.15	.21	1.76	.12	.11	.14	1.76	.09	.08	.12
2	1.84	1.72	.32	.17	.34	1.80	.13	.13	.13	1.76	.06	.09	.10
3	1.69	1.67	.15	.16	.15	1.66	.13	.11	.13	1.63	.10	.08	.11
4	1.62	1.50	.10	.13	.15	1.51	.06	.09	.12	1.53	.07	.07	.11
5	1.59	1.37	.11	.12	.24	1.47	.08	.09	.14	1.54	.06	.07	.07
6	2.60	2.50	.19	.33	.21	2.56	.10	.25	.11	2.60	.16	.16	.16
7	2.13	1.82	.13	.17	.36	1.92	.13	.12	.24	2.00	.09	.09	.15
8	2.02	1.92	.21	.17	.23	1.92	.12	.12	.15	1.99	.06	.09	.06
9	1.97	1.74	.12	.15	.25	1.88	.09	.11	.12	1.86	.06	.08	.12
10	1.90	1.76	.17	.17	.22	1.84	.17	.12	.18	1.89	.15	.09	.15
11	1.28	1.38	.11	.12	.14	1.42	.02	.09	.14	1.43	.03	.06	.15
12	1.27	1.33	.09	.12	.10	1.38	.07	.08	.13	1.42	.05	.06	.15
13	1.26	1.25	.09	.12	.09	1.36	.10	.09	.14	1.38	.07	.06	.13
14	1.26	1.27	.11	.11	.11	1.32	.06	.08	.08	1.29	.06	.06	.06
15	1.24	1.17	.14	.11	.15	1.21	.07	.08	.07	1.25	.03	.06	.03
total	1.66	1.61	.14	.15	.14	1.67	.09	.11	.09	1.69	.07	.08	.07

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E1d: i=15; α, large *e*, large *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	1.84	1.65	.08	.15	.20	1.78	.10	.11	.11	1.77	.07	.08	.09
2	1.84	1.75	.16	.17	.18	1.77	.11	.12	.13	1.76	.04	.09	.08
3	1.69	1.66	.20	.16	.20	1.66	.12	.12	.12	1.63	.09	.08	.10
4	1.62	1.51	.12	.13	.16	1.52	.08	.09	.12	1.55	.07	.07	.09
5	1.59	1.44	.11	.13	.18	1.49	.09	.09	.13	1.55	.05	.07	.06
6	2.60	2.42	.11	.31	.21	2.52	.13	.23	.15	2.56	.13	.16	.13
7	2.13	1.83	.11	.17	.31	1.96	.10	.13	.19	2.01	.08	.09	.14
8	2.02	1.85	.25	.17	.30	1.88	.08	.12	.16	1.96	.07	.09	.09
9	1.97	1.72	.16	.15	.29	1.86	.12	.11	.16	1.87	.05	.08	.11
10	1.90	1.79	.11	.17	.15	1.86	.11	.12	.11	1.90	.13	.09	.13
11	1.28	1.42	.09	.12	.16	1.44	.06	.09	.17	1.44	.06	.06	.17
12	1.27	1.30	.08	.11	.08	1.35	.05	.08	.09	1.40	.05	.06	.13
13	1.26	1.23	.11	.11	.11	1.35	.11	.09	.14	1.37	.07	.06	.13
14	1.26	1.33	.10	.12	.12	1.34	.08	.08	.11	1.29	.06	.05	.06
15	1.24	1.15	.10	.11	.13	1.24	.06	.08	.06	1.26	.04	.06	.04
total	1.66	1.60	.12	.15	.13	1.67	.09	.11	.10	1.69	.07	.08	.09

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E1e: i=30; α, small *e*, small *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	1.84	1.64	.18	.14	.27	1.73	.13	.10	.17	1.76	.11	.07	.13
2	1.84	1.64	.18	.15	.27	1.72	.09	.11	.15	1.80	.05	.08	.06
3	1.69	1.50	.13	.14	.23	1.55	.09	.10	.16	1.62	.04	.07	.08
4	1.62	1.35	.13	.12	.29	1.42	.03	.08	.20	1.49	.04	.06	.13
5	1.59	1.36	.10	.12	.25	1.47	.11	.09	.16	1.51	.06	.07	.10
6	2.60	2.39	.36	.26	.41	2.44	.27	.19	.31	2.53	.13	.14	.14
7	2.13	1.90	.16	.17	.28	1.97	.09	.12	.18	2.03	.08	.09	.12
8	2.02	1.74	.10	.15	.29	1.88	.10	.11	.17	1.93	.06	.08	.10
9	1.97	1.79	.16	.15	.24	1.84	.16	.11	.20	1.89	.07	.08	.10
10	1.90	1.65	.13	.15	.28	1.75	.07	.11	.16	1.84	.06	.08	.08
11	1.28	1.14	.11	.10	.17	1.19	.07	.07	.11	1.22	.04	.05	.07
12	1.27	1.07	.10	.09	.22	1.15	.03	.07	.12	1.20	.04	.05	.08
13	1.26	1.12	.06	.10	.15	1.20	.04	.07	.07	1.22	.01	.05	.04
14	1.26	1.09	.04	.10	.17	1.19	.05	.07	.08	1.20	.04	.05	.07
15	1.24	1.11	.11	.10	.17	1.16	.08	.07	.11	1.19	.03	.05	.05
16	1.47	1.29	.07	.11	.19	1.38	.09	.08	.12	1.39	.03	.06	.08
17	1.44	1.28	.12	.11	.20	1.33	.08	.08	.13	1.38	.05	.06	.07
18	1.42	1.33	.16	.12	.18	1.32	.08	.08	.12	1.35	.05	.06	.08
19	1.32	1.20	.08	.10	.14	1.23	.07	.07	.11	1.27	.07	.05	.08
20	1.29	1.15	.07	.10	.15	1.20	.05	.07	.10	1.23	.04	.05	.07
21	1.06	.99	.02	.09	.07	1.02	.04	.07	.05	1.02	.03	.05	.05
22	1.05	.97	.09	.09	.12	.97	.05	.06	.08	1.01	.04	.04	.05
23	1.01	.89	.05	.08	.13	.91	.03	.06	.10	.95	.03	.04	.06

24	.97	.89	.11	.08	.13	.88	.06	.06	.10	.91	.05	.04	.07
25	.86	.77	.10	.07	.13	.81	.04	.05	.06	.82	.03	.04	.05
26	1.23	1.04	.09	.09	.21	1.12	.08	.07	.13	1.80	.06	.05	.57
27	1.2	1.05	.06	.10	.16	1.10	.06	.07	.11	1.12	.02	.05	.08
28	1.19	1.08	.11	.10	.15	1.15	.07	.07	.08	1.12	.02	.05	.07
29	1.15	1.00	.06	.09	.16	1.03	.04	.06	.12	1.08	.05	.05	.08
30	1.08	1.00	.15	.09	.17	1.03	.08	.06	.09	1.04	.04	.04	.05
total	1.42	1.28	.11	.12	.17	1.34	.07	.09	.10	1.38	.04	.06	.05

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E1f: i=30; α, small *e*, large *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	1.84	1.61	.16	.14	.28	1.71	.12	.10	.17	1.70	.09	.07	.16
2	1.84	1.59	.14	.14	.28	1.68	.07	.11	.17	1.74	.01	.08	.10
3	1.69	1.48	.13	.14	.24	1.54	.10	.10	.18	1.61	.03	.07	.08
4	1.62	1.37	.14	.12	.28	1.43	.05	.09	.19	1.48	.06	.06	.15
5	1.59	1.35	.09	.12	.25	1.46	.10	.09	.16	1.50	.07	.06	.11
6	2.60	2.41	.39	.27	.43	2.47	.28	.20	.30	2.54	.15	.14	.16
7	2.13	1.90	.17	.17	.28	1.96	.13	.12	.21	2.05	.12	.09	.14
8	2.02	1.77	.15	.15	.29	1.87	.12	.11	.19	1.91	.06	.08	.12
9	1.97	1.80	.18	.15	.22	1.82	.18	.11	.23	1.87	.07	.08	.12
10	1.90	1.65	.13	.15	.28	1.75	.06	.11	.16	1.83	.02	.08	.07
11	1.28	1.15	.13	.10	.18	1.19	.07	.07	.11	1.22	.02	.05	.06
12	1.27	1.09	.11	.09	.21	1.15	.06	.07	.13	1.17	.03	.05	.10
13	1.26	1.13	.08	.10	.15	1.21	.06	.07	.07	1.25	.02	.05	.02
14	1.26	1.09	.05	.10	.17	1.19	.04	.07	.08	1.19	.03	.05	.07
15	1.24	1.10	.10	.09	.17	1.16	.07	.07	.10	1.20	.05	.05	.06
16	1.47	1.30	.11	.11	.20	1.37	.13	.08	.16	1.38	.04	.06	.09
17	1.44	1.29	.11	.11	.18	1.35	.07	.08	.11	1.40	.04	.06	.05
18	1.42	1.31	.14	.12	.17	1.31	.08	.08	.13	1.31	.02	.06	.11
19	1.32	1.20	.09	.10	.15	1.23	.06	.08	.10	1.23	.02	.05	.09
20	1.29	1.14	.06	.10	.16	1.19	.05	.07	.11	1.20	.02	.05	.09
21	1.06	.96	.02	.09	.10	1.00	.03	.07	.06	1.00	.03	.05	.06
22	1.05	.95	.09	.09	.13	.95	.05	.06	.11	.99	.03	.04	.06
23	1.01	.90	.04	.08	.11	.92	.03	.06	.09	.95	.04	.04	.07

24	.97	.89	.09	.08	.12	.89	.06	.06	.10	.89	.03	.04	.08
25	.86	.77	.09	.07	.12	.82	.04	.05	.05	.83	.04	.04	.05
26	1.23	1.04	.09	.09	.21	1.11	.07	.07	.13	1.19	.06	.05	.07
27	1.2	1.04	.07	.10	.17	1.09	.07	.07	.13	1.10	.03	.05	.10
28	1.19	1.08	.09	.10	.14	1.14	.06	.07	.07	1.13	.01	.05	.06
29	1.15	.99	.07	.09	.17	1.03	.04	.06	.12	1.10	.05	.05	.07
30	1.08	1.02	.15	.09	.16	1.03	.07	.06	.08	1.03	.02	.04	.05
total	1.42	1.27	.11	.11	.18	1.34	.08	.08	.11	1.37	.04	.06	.06

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E1g: i=30; α, large *e*, small *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	1.84	1.65	.12	.14	.22	1.78	.10	.10	.11	1.76	.08	.07	.11
2	1.84	1.58	.14	.15	.29	1.69	.10	.11	.18	1.76	.06	.08	.10
3	1.69	1.47	.10	.13	.24	1.54	.06	.10	.16	1.61	.05	.07	.09
4	1.62	1.33	.16	.11	.33	1.44	.08	.08	.19	1.49	.07	.06	.14
5	1.59	1.37	.19	.12	.29	1.49	.18	.09	.20	1.52	.11	.06	.13
6	2.60	2.36	.33	.29	.40	2.36	.25	.20	.34	2.49	.23	.14	.25
7	2.13	1.87	.19	.16	.32	1.92	.12	.11	.24	2.01	.09	.09	.15
8	2.02	1.74	.13	.14	.30	1.89	.12	.11	.17	1.91	.05	.08	.12
9	1.97	1.75	.16	.14	.27	1.79	.13	.10	.22	1.88	.06	.08	.10
10	1.90	1.66	.13	.15	.27	1.78	.12	.13	.16	1.85	.07	.08	.08
11	1.28	1.20	.08	.10	.11	1.21	.06	.07	.09	1.22	.04	.05	.07
12	1.27	1.07	.10	.09	.21	1.16	.04	.07	.11	1.21	.04	.05	.07
13	1.26	1.13	.08	.10	.15	1.19	.04	.07	.08	1.21	.03	.05	.05
14	1.26	1.10	.07	.10	.17	1.19	.03	.07	.07	1.19	.02	.05	.07
15	1.24	1.09	.12	.09	.19	1.15	.03	.07	.09	1.19	.03	.05	.05
16	1.47	1.31	.07	.11	.17	1.41	.10	.08	.11	1.41	.04	.06	.07
17	1.44	1.34	.15	.11	.18	1.36	.08	.08	.11	1.42	.07	.06	.07
18	1.42	1.33	.18	.12	.20	1.35	.08	.09	.10	1.35	.07	.06	.09
19	1.32	1.17	.05	.10	.15	1.23	.09	.07	.12	1.26	.09	.05	.10
20	1.29	1.09	.10	.10	.22	1.18	.05	.07	.12	1.22	.07	.05	.09
21	1.06	.99	.05	.09	.08	1.05	.05	.06	.05	1.03	.02	.04	.03
22	1.05	.97	.08	.08	.11	.97	.07	.06	.10	1.01	.05	.04	.06
23	1.01	.90	.03	.08	.11	.90	.03	.05	.11	.95	.04	.04	.07

24	.97	.86	.09	.08	.14	.89	.04	.05	.11	.90	.04	.04	.08
25	.86	.76	.06	.07	.11	.80	.05	.05	.07	.81	.03	.03	.05
26	1.23	1.05	.08	.09	.19	1.12	.08	.07	.13	1.18	.04	.05	.06
27	1.2	1.06	.06	.09	.15	1.09	.05	.06	.12	1.12	.02	.05	.08
28	1.19	1.07	.02	.09	.13	1.13	.05	.07	.07	1.11	.02	.05	.08
29	1.15	1.02	.07	.09	.14	1.06	.02	.06	.09	1.09	.04	.04	.07
30	1.08	1.05	.13	.09	.13	1.05	.06	.06	.06	1.05	.04	.04	.05
total	1.42	1.28	.11	.11	.17	1.34	.09	.08	.12	1.37	.05	.06	.07

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E1h: i=30; α, large *e*, large *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	1.84	1.68	.14	.14	.21	1.76	.12	.11	.14	1.79	.12	.08	.13
2	1.84	1.61	.15	.15	.27	1.68	.12	.11	.20	1.77	.06	.08	.09
3	1.69	1.43	.09	.13	.27	1.48	.05	.10	.21	1.60	.05	.07	.10
4	1.62	1.29	.13	.11	.35	1.39	.07	.08	.24	1.50	.05	.06	.13
5	1.59	1.35	.13	.11	.27	1.46	.08	.09	.15	1.51	.07	.06	.10
6	2.60	2.38	.41	.29	.48	2.55	.32	.21	.32	2.54	.18	.15	.18
7	2.13	1.88	.19	.16	.31	1.98	.08	.12	.17	2.02	.09	.08	.14
8	2.02	1.76	.12	.15	.28	1.90	.07	.11	.13	1.93	.04	.08	.09
9	1.97	1.72	.16	.14	.29	1.76	.09	.10	.22	1.88	.08	.08	.12
10	1.90	1.65	.08	.14	.26	1.77	.08	.11	.15	1.86	.08	.08	.08
11	1.28	1.20	.10	.10	.12	1.19	.05	.07	.10	1.23	.02	.05	.05
12	1.27	1.07	.12	.09	.23	1.15	.03	.07	.12	1.22	.04	.05	.06
13	1.26	1.14	.10	.10	.15	1.23	.05	.07	.05	1.22	.04	.05	.05
14	1.26	1.10	.06	.10	.17	1.21	.04	.07	.06	1.21	.03	.05	.05
15	1.24	1.09	.13	.09	.19	1.17	.07	.07	.09	1.20	.03	.05	.05
16	1.47	1.32	.07	.11	.16	1.39	.10	.08	.12	1.43	.04	.06	.05
17	1.44	1.35	.12	.11	.15	1.39	.08	.08	.09	1.42	.05	.06	.05
18	1.42	1.34	.20	.12	.21	1.34	.07	.09	.10	1.36	.06	.06	.08
19	1.32	1.15	.07	.10	.18	1.20	.09	.07	.15	1.26	.10	.05	.11
20	1.29	1.10	.06	.10	.19	1.17	.07	.07	.13	1.24	.07	.05	.08
21	1.06	.99	.04	.09	.08	1.02	.04	.06	.05	1.02	.02	.04	.04
22	1.05	.96	.07	.08	.11	1.00	.05	.06	.07	1.01	.04	.04	.05
23	1.01	.88	.03	.07	.13	.92	.04	.06	.09	.94	.04	.04	.08

24	.97	.86	.06	.07	.12	.88	.04	.05	.09	.91	.04	.04	.07
25	.86	.77	.06	.07	.13	.78	.06	.05	.10	.82	.03	.03	.05
26	1.23	1.07	.10	.09	.18	1.16	.10	.07	.12	1.19	.04	.05	.05
27	1.2	1.05	.06	.09	.16	1.10	.03	.06	.10	1.12	.03	.05	.08
28	1.19	1.05	.02	.09	.14	1.11	.04	.07	.08	1.11	.03	.05	.08
29	1.15	1.04	.06	.09	.12	1.06	.05	.06	.10	1.09	.04	.04	.07
30	1.08	1.04	.10	.09	.10	1.05	.04	.06	.05	1.06	.03	.04	.03
total	1.42	1.28	.10	.11	.17	1.34	.08	.08	.11	1.38	.06	.06	.07

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E2a: i=15; δ_1, small *e*, small *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	-1.22	-1.21	.06	.10	.06	-1.29	.07	.07	.09	-1.25	.02	.05	.03
2	-1.35	-1.22	.12	.11	.17	-1.40	.05	.08	.07	-1.37	.02	.05	.02
3	.20	.42	.14	.09	.26	.23	.07	.06	.07	.21	.05	.04	.05
4	-1.43	-1.29	.20	.12	.24	-1.53	.04	.08	.12	-1.50	.05	.06	.08
5	-2.41	-2.11	.39	.24	.49	-2.55	.18	.15	.22	-2.49	.08	.11	.11
6	-1.95	-1.89	.21	.14	.21	-2.01	.03	.09	.10	-1.98	.02	.06	.09
7	-2.46	-2.37	.20	.20	.21	-2.54	.11	.13	.13	-2.50	.09	.09	.09
8	-.66	-.60	.08	.09	.10	-.68	.06	.06	.06	-.67	.06	.04	.06
9	-.94	-.93	.07	.09	.07	-.96	.03	.06	.03	-.97	.04	.04	.05
10	-2.81	-2.85	.24	.28	.24	-2.91	.22	.17	.24	-2.84	.16	.12	.16
11	-1.72	-1.72	.13	.14	.13	-1.78	.05	.09	.07	-1.75	.03	.07	.04
12	-1.5	-1.27	.15	.13	.27	-1.56	.12	.09	.13	-1.55	.09	.06	.10
13	.43	.75	.19	.14	.37	.46	.11	.09	.11	.44	.10	.06	.10
14	-1.38	-1.27	.24	.14	.26	-1.39	.09	.08	.09	-1.37	.07	.06	.07
15	-.27	-.08	.14	.14	.23	-.30	.05	.09	.05	-.31	.05	.06	.06
total	-1.29	-1.17	.12	.14	.16	-1.35	.07	.09	.09	-1.33	.05	.06	.06

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E2b: i=15; δ_1, small *e*, large *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	-1.22	-1.37	.15	.11	.21	-1.26	.11	.07	.11	-1.25	.03	.05	.04
2	-1.35	-1.48	.11	.11	.17	-1.40	.13	.08	.13	-1.36	.03	.05	.03
3	.20	.254	.13	.10	.14	.25	.11	.06	.12	.22	.06	.04	.06
4	-1.43	-1.54	.16	.12	.21	-1.52	.05	.08	.10	-1.50	.04	.06	.08
5	-2.41	-2.63	.38	.23	.43	-2.52	.07	.14	.13	-2.46	.09	.10	.10
6	-1.95	-2.04	.14	.13	.16	-2.00	.15	.09	.15	-1.98	.03	.06	.04
7	-2.46	-2.63	.20	.20	.26	-2.51	.12	.13	.13	-2.50	.11	.09	.11
8	-.66	-.739	.08	.09	.11	-.68	.10	.06	.10	-.67	.05	.04	.05
9	-.94	-1.02	.06	.09	.10	-.95	.08	.06	.08	-.98	.04	.04	.05
10	-2.81	-3.13	.21	.26	.38	-2.92	.16	.16	.19	-2.85	.16	.11	.16
11	-1.72	-1.88	.15	.14	.21	-1.77	.07	.09	.08	-1.74	.04	.07	.04
12	-1.5	-1.58	.13	.13	.15	-1.55	.08	.09	.09	-1.56	.10	.06	.11
13	.43	.519	.15	.14	.17	.50	.10	.09	.12	.45	.08	.06	.08
14	-1.38	-1.44	.21	.12	.21	-1.39	.04	.09	.04	-1.39	.06	.06	.06
15	-.27	-.235	.07	.14	.07	-.26	.04	.09	.04	-.29	.03	.06	.03
total	-1.29	-1.39	.13	.14	.16	-1.33	.09	.09	.09	-1.32	.06	.06	.06

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E2c: i=15; δ_1, large *e*, small *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	-1.22	-1.34	.15	.09	.19	-1.28	.04	.06	.07	-1.26	.02	.04	.04
2	-1.35	-1.49	.12	.10	.18	-1.40	.03	.06	.05	-1.38	.01	.04	.03
3	.20	.19	.06	.07	.06	.23	.07	.05	.07	.21	.05	.03	.05
4	-1.43	-1.54	.15	.10	.18	-1.52	.05	.07	.10	-1.49	.03	.05	.06
5	-2.41	-2.73	.33	.22	.45	-2.52	.13	.13	.17	-2.45	.08	.09	.08
6	-1.95	-2.08	.15	.13	.19	-2.00	.04	.08	.11	-1.97	.03	.05	.03
7	-2.46	-2.72	.21	.19	.27	-2.55	.10	.12	.13	-2.50	.10	.08	.10
8	-.66	-.74	.09	.07	.12	-.69	.04	.05	.05	-.68	.04	.03	.04
9	-.94	-1.01	.08	.08	.10	-.96	.04	.05	.04	-.97	.03	.03	.04
10	-2.81	-3.19	.47	.30	.36	-2.89	.23	.17	.24	-2.83	.20	.11	.20
11	-1.72	-1.89	.16	.12	.23	-1.79	.05	.08	.08	-1.77	.03	.06	.05
12	-1.5	-1.62	.10	.11	.15	-1.56	.08	.08	.10	-1.54	.06	.05	.07
13	.43	.51	.05	.12	.09	.44	.06	.07	.06	.43	.04	.05	.04
14	-1.38	-1.47	.17	.11	.19	-1.42	.04	.07	.05	-1.40	.03	.05	.03
15	-.27	-.29	.07	.11	.07	-.28	.08	.07	.08	-.31	.04	.05	.05
total	-1.29	-1.43	.13	.13	.19	-1.35	.09	.08	.10	-1.33	.06	.05	.07

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E2d: i=15; δ_1, large *e*, large *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	-1.22	-1.34	.11	.09	.16	-1.27	.06	.06	.07	-1.26	.03	.04	.05
2	-1.35	-1.45	.10	.09	.14	-1.39	.03	.06	.05	-1.36	.01	.04	.01
3	.20	.20	.12	.07	.12	.22	.08	.05	.08	.20	.06	.03	.06
4	-1.43	-1.53	.15	.10	.18	-1.51	.04	.07	.08	-1.49	.02	.04	.06
5	-2.41	-2.69	.29	.20	.40	-2.52	.13	.12	.17	-2.47	.06	.08	.08
6	-1.95	-2.02	.10	.11	.12	-1.99	.05	.07	.06	-1.97	.03	.05	.03
7	-2.46	-2.62	.19	.17	.24	-2.52	.10	.11	.11	-2.51	.09	.08	.10
8	-.66	-.73	.06	.07	.09	-.68	.05	.05	.05	-.67	.04	.03	.04
9	-.94	-.99	.04	.07	.06	-.95	.04	.05	.04	-.97	.04	.03	.05
10	-2.81	-3.14	.43	.28	.54	-2.88	.21	.15	.22	-2.82	.17	.10	.17
11	-1.72	-1.87	.16	.12	.21	-1.79	.05	.08	.08	-1.78	.03	.05	.06
12	-1.5	-1.59	.15	.11	.17	-1.58	.09	.07	.12	-1.55	.06	.05	.07
13	.43	.56	.12	.12	.17	.43	.06	.07	.06	.43	.06	.05	.06
14	-1.38	-1.46	.17	.10	.18	-1.42	.04	.06	.05	-1.41	.03	.05	.04
15	-.27	-.26	.05	.11	.05	-.28	.04	.07	.04	-.30	.05	.05	.05
total	-1.29	-1.39	.14	.12	.17	-1.34	.08	.08	.09	-1.33	.06	.05	.07

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E2e: i=30; δ_1, small *e*, small *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	-1.22	-1.33	.11	.07	.15	-1.26	.02	.07	.04	-1.23	.03	.04	.03
2	-1.35	-1.48	.11	.07	.17	-1.41	.04	.07	.07	-1.39	.03	.05	.05
3	.20	.22	.13	.07	.11	.19	.07	.06	.07	.19	.06	.04	.06
4	-1.43	-1.67	.10	.08	.26	-1.55	.04	.08	.12	-1.48	.04	.06	.06
5	-2.41	-2.69	.21	.14	.35	-2.53	.04	.15	.12	-2.53	.10	.10	.15
6	-1.95	-2.19	.12	.09	.26	-2.10	.05	.09	.07	-2.03	.10	.06	.12
7	-2.46	-2.66	.26	.13	.32	-2.55	.15	.13	.17	-2.50	.04	.09	.05
8	-.66	-.72	.09	.06	.10	-.69	.05	.06	.05	-.68	.02	.04	.02
9	-.94	-1.05	.11	.06	.15	-1.03	.07	.06	.11	-.99	.02	.04	.05
10	-2.81	-3.06	.40	.17	.47	-2.94	.23	.17	.26	-2.88	.16	.12	.17
11	-2.02	-2.34	.10	.12	.33	-2.21	.11	.12	.21	-2.15	.14	.08	.19
12	-.96	-1.01	.11	.09	.12	-1.06	.05	.09	.11	-1.00	.02	.06	.04
13	-1.12	-1.20	.06	.08	.10	-1.17	.03	.08	.05	-1.13	.04	.05	.04
14	-2.38	-2.82	.19	.14	.47	-2.56	.10	.14	.20	-2.49	.13	.10	.17
15	-.72	-.75	.10	.10	.10	-.75	.06	.10	.06	-.76	.07	.06	.08
16	-1.72	-1.86	.07	.10	.15	-1.79	.11	.10	.13	-1.79	.07	.07	.09
17	-1.5	-1.64	.14	.09	.19	-1.57	.08	.09	.10	-1.53	.04	.06	.05
18	.43	.45	.17	.09	.17	.45	.08	.09	.08	.43	.06	.06	.06
19	-1.38	-1.59	.12	.09	.24	-1.52	.11	.09	.17	-1.46	.07	.06	.10
20	-.27	-.29	.11	.09	.11	-.25	.09	.09	.09	-.28	.06	.06	.06
21	.19	.07	.08	.08	.13	.15	.03	.08	.05	.16	.02	.06	.03
22	-1.66	-1.65	.11	.13	.11	-1.69	.05	.13	.05	-1.71	.06	.09	.07
23	-1.41	-1.53	.15	.11	.19	-1.55	.08	.11	.16	-1.49	.04	.07	.08

24	-1.12	-1.25	.15	.13	.19	-1.17	.13	.14	.13	-1.17	.10	.09	.11
25	-1.94	-2.10	.21	.15	.26	-2.05	.10	.15	.14	-2.03	.07	.11	.11
26	-1.02	-1.14	.10	.08	.15	-1.08	.08	.08	.10	-1.01	.05	.05	.05
27	-1.26	-1.35	.33	.13	.34	-1.33	.14	.14	.15	-1.27	.06	.09	.06
28	.21	.264	.12	.08	.13	.187	.04	.08	.04	.22	.04	.05	.04
29	-1.57	-1.75	.20	.12	26	-1.69	.12	.12	.16	-1.62	.09	.08	.10
30	-1.55	-1.66	.14	.11	.17	-1.67	.12	.11	.16	-1.60	.04	.07	.06
total	-1.26	-1.39	.14	.10	.19	-1.34	.10	.10	.12	-1.31	.08	.07	.09

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E2f: i=30; δ_1, small *e*, large *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	-1.22	-1.32	.11	.10	.14	-1.26	.05	.07	.06	-1.25	.02	.07	.03
2	-1.35	-1.45	.11	.11	.14	-1.39	.04	.07	.05	-1.39	.03	.08	.05
3	.20	.26	.15	.10	.16	.21	.08	.07	.08	.20	.06	.07	.06
4	-1.43	-1.64	.09	.13	.22	-1.55	.04	.08	.12	-1.48	.04	.06	.06
5	-2.41	-2.69	.18	.22	.33	-2.53	.07	.14	.13	-2.48	.06	.06	.09
6	-1.95	-2.19	.15	.14	.28	-2.09	.06	.09	.15	-2.02	.06	.14	.09
7	-2.46	-2.65	.22	.19	.29	-2.55	.13	.13	.15	-2.44	.12	.09	.12
8	-.66	-.73	.11	.09	.13	-.70	.07	.06	.08	-.69	.01	.08	.03
9	-.94	-1.05	.12	.09	.16	-1.03	.08	.06	.12	-1.02	.03	.08	.08
10	-2.81	-3.07	.41	.25	.48	-2.97	.23	.17	.28	-2.95	.11	.08	.17
11	-2.02	-2.31	.11	.18	.31	-2.20	.09	.12	.20	-2.13	.05	.05	.12
12	-.96	-1.02	.12	.13	.13	-1.06	.08	.09	.12	-.99	.04	.05	.05
13	-1.12	-1.22	.06	.12	.11	-1.18	.03	.08	.06	-1.09	.04	.05	.05
14	-2.38	-2.84	.19	.23	.49	-2.56	.11	.14	.21	-2.47	.11	.05	.14
15	-.72	-.71	.11	.14	.11	-.75	.06	.10	.06	-.72	.08	.05	.08
16	-1.72	-1.87	.10	.14	.18	-1.81	.11	.10	.14	-1.77	.07	.06	.08
17	-1.5	-1.63	.12	.13	.17	-1.56	.07	.09	.09	-1.51	.02	.06	.02
18	.43	.50	.18	.14	.19	.48	.12	.09	.13	.49	.07	.06	.09
19	-1.38	-1.62	.13	.13	.27	-1.54	.09	.09	.18	-1.46	.04	.05	.08
20	-.27	-.26	.08	.14	.08	-.25	.07	.09	.07	-.25	.05	.05	.05
21	.19	.06	.11	.12	.17	.14	.08	.08	.09	.12	.02	.05	.07
22	-1.66	-1.72	.14	.18	.15	-1.71	.06	.13	.07	-1.77	.02	.04	.11
23	-1.41	-1.52	.15	.15	.18	-1.54	.10	.11	.16	-1.45	.07	.04	.08

24	-1.12	-1.26	.13	.19	.19	-1.21	.14	.13	.16	-1.22	.08	.04	.12
25	-1.94	-2.09	.18	.23	.23	-2.04	.11	.15	.14	-2.00	.04	.04	.07
26	-1.02	-1.16	.11	.12	.17	-1.09	.09	.08	.11	-1.02	.05	.05	.05
27	-1.26	-1.29	.28	.20	.28	-1.29	.15	.13	.15	-1.23	.04	.05	.05
28	.21	.25	.13	.12	.13	.18	.08	.08	.08	.24	.02	.05	.03
29	-1.57	-1.77	.24	.18	.31	-1.70	.11	.12	.17	-1.59	.10	.05	.10
30	-1.55	-1.61	.12	.15	.13	-1.66	.11	.11	.15	-1.62	.04	.04	.08
total	-1.26	-1.39	.14	.15	.19	-1.34	.10	.10	.12	-1.30	.06	.07	.07

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E2g: i=30; δ_1, large *e*, small *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	-1.22	-1.35	.12	.09	.17	-1.27	.06	.06	.07	-1.25	.02	.04	.03
2	-1.35	-1.52	.12	.10	.20	-1.41	.04	.07	.07	-1.39	.02	.04	.04
3	.20	.22	.11	.08	.11	.23	.07	.05	.07	.20	.04	.03	.04
4	-1.43	-1.68	.11	.11	.27	-1.57	.05	.07	.14	-1.49	.02	.04	.06
5	-2.41	-2.66	.13	.19	.28	-2.49	.05	.13	.09	-2.52	.09	.09	.14
6	-1.95	-2.19	.16	.12	.28	-2.14	.09	.09	.21	-2.03	.06	.05	.10
7	-2.46	-2.69	.29	.19	.37	-2.63	.06	.13	.18	-2.52	.12	.08	.13
8	-.66	-.72	.08	.07	.10	-.69	.02	.05	.03	-.69	.01	.05	.03
9	-.94	-1.06	.12	.08	.16	-1.02	.08	.05	.11	-.99	.04	.03	.06
10	-2.81	-3.07	.30	.25	.39	-2.85	.18	.15	.18	-2.83	.09	.10	.09
11	-2.02	-2.29	.16	.15	.31	-2.18	.16	.10	.22	-2.11	.12	.07	.15
12	-.96	-1.05	.10	.10	.13	-.99	.05	.07	.05	-.98	.03	.04	.03
13	-1.12	-1.21	.07	.09	.11	-1.16	.05	.07	.06	-1.13	.06	.04	.06
14	-2.38	-2.82	.15	.21	.46	-2.54	.08	.13	.17	-2.53	.05	.09	.15
15	-.72	-.79	.08	.11	.10	-.75	.03	.08	.04	-.74	.02	.05	.02
16	-1.72	-1.91	.09	.12	.21	-1.79	.04	.08	.08	-1.81	.05	.06	.10
17	-1.50	-1.66	.08	.11	.17	-1.58	.05	.07	.09	-1.54	.05	.05	.06
18	.43	.47	.15	.11	.15	.47	.11	.08	.11	.45	.07	.05	.07
19	-1.38	-1.61	.12	.11	.25	-1.51	.07	.08	.14	-1.46	.04	.05	.08
20	-.27	-.27	.10	.12	.10	-.27	.08	.08	.08	-.27	.04	.05	.04
21	.19	.08	.08	.09	.13	.16	.05	.06	.05	.16	.03	.04	.04
22	-1.66	-1.74	.11	.14	.13	-1.69	.04	.10	.05	-1.72	.04	.07	.07
23	-1.41	-1.59	.09	.13	.20	-1.55	.04	.09	.14	-1.49	.05	.06	.09

24	-1.12	-1.27	.11	.15	.18	-1.20	.12	.11	.14	-1.17	.05	.07	.07
25	-1.94	-2.27	.21	.19	.39	-2.09	.09	.12	.17	-2.05	.07	.09	.13
26	-1.02	-1.17	.11	.10	.18	-1.08	.09	.07	.10	-1.03	.05	.04	.05
27	-1.26	-1.39	.26	.16	.29	-1.36	.07	.11	.12	-1.30	.05	.07	.06
28	.21	.22	.10	.09	.10	.21	.03	.06	.03	.22	.03	.04	.03
29	-1.57	-1.75	.13	.14	.22	-1.68	.09	.09	.14	-1.62	.06	.06	.07
30	-1.55	-1.69	.12	.12	.18	-1.65	.08	.09	.12	-1.61	.04	.06	.07
total	-1.26	-1.41	.13	.13	.19	-1.34	.08	.09	.11	1.29	.05	.06	.05

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E2h: i=30; δ_1, large *e*, large *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	-1.22	-1.34	.09	.08	.15	-1.28	.03	.06	.06	-1.24	.02	.04	.02
2	-1.35	-1.49	.10	.09	.17	-1.42	.05	.06	.08	-1.39	.02	.04	.04
3	.20	.24	.10	.08	.10	.24	.08	.05	.08	.19	.04	.03	.04
4	-1.43	-1.65	.11	.11	.24	-1.53	.07	.07	.12	-1.47	.03	.04	.05
5	-2.41	-2.67	.11	.19	.28	-2.54	.04	.12	.13	-2.52	.07	.08	.13
6	-1.95	-2.18	.20	.12	.30	-2.08	.06	.08	.14	-2.02	.06	.05	.09
7	-2.46	-2.66	.23	.17	.30	-2.57	.09	.11	.14	-2.52	.12	.08	.13
8	-.66	-.73	.09	.07	.11	-.69	.05	.05	.05	-.67	.01	.03	.01
9	-.94	-1.05	.11	.07	.15	-1.02	.06	.05	.10	-.98	.02	.03	.04
10	-2.81	-3.07	.31	.23	.40	-2.92	.17	.15	.20	-2.85	.10	.10	.10
11	-2.02	-2.29	.18	.15	.32	-2.19	.09	.10	.19	-2.11	.11	.07	.14
12	-.96	-1.01	.06	.10	.07	-1.02	.04	.06	.07	-.99	.03	.04	.04
13	-1.12	-1.21	.06	.09	.10	-1.16	.02	.06	.04	-1.13	.04	.04	.04
14	-2.38	-2.82	.19	.21	.46	-2.52	.11	.12	.17	-2.49	.07	.09	.13
15	-.72	-.71	.14	.11	.14	-.73	.07	.07	.07	-.74	.04	.05	.04
16	-1.72	-1.8	.09	.11	.12	-1.81	.07	.08	.11	-1.77	.04	.05	.06
17	-1.5	-1.62	.08	.10	.14	-1.56	.04	.07	.07	-1.52	.02	.05	.02
18	.43	.47	.18	.11	.18	.48	.10	.08	.11	.46	.06	.05	.06
19	-1.38	-1.55	.12	.11	.20	-1.49	.07	.07	.13	-1.44	.02	.05	.06
20	-.27	-.26	.08	.12	.08	-.19	.12	.08	.14	-.26	.04	.05	.04
21	.19	.09	.10	.09	.14	.15	.08	.06	.08	.15	.02	.04	.04
22	-1.66	-1.75	.11	.14	.14	-1.68	.03	.09	.03	-1.70	.03	.06	.05
23	-1.41	-1.56	.10	.12	.18	-1.53	.07	.08	.13	-1.48	.04	.06	.08

24	-1.12	-1.19	.07	.15	.09	-1.19	.06	.10	.09	-1.15	.03	.07	.04
25	-1.94	-2.15	.13	.18	.24	-2.04	.04	.12	.10	-2.02	.07	.08	.10
26	-1.02	-1.15	.10	.10	.16	-1.08	.07	.06	.09	-1.02	.04	.04	.04
27	-1.26	-1.27	.25	.16	.25	-1.33	.05	.10	.08	-1.26	.06	.07	.06
28	.21	.22	.10	.09	.10	.19	.03	.06	.03	.21	.03	.04	.03
29	-1.57	-1.76	.11	.13	.21	-1.70	.07	.09	.14	-1.62	.06	.06	.07
30	-1.55	-1.65	.10	.12	.14	-1.66	.07	.08	.13	-1.59	.04	.06	.05
total	-1.26	-1.39	.14	.12	.19	-1.33	.08 *	.08	.10	-1.30	.05	.05	.06

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E3a: i=15; δ_2, small *e*, small *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	-.10	-.09	.12	.10	.12	-.09	.02	.06	.02	-.09	.01	.04	.01
2	-.56	-.55	.08	.12	.08	-.55	.06	.08	.06	-.55	.06	.05	.06
3	.69	.65	.08	.11	.08	.66	.05	.07	.06	.67	.05	.05	.05
4	-.14	-.17	.16	.10	.16	-.16	.09	.07	.09	-.14	.07	.05	.07
5	-1.17	-1.20	.09	.11	.09	-1.21	.07	.08	.08	-1.18	.05	.05	.05
6	.01	.05	.27	.18	.27	.03	.10	.12	.10	.02	.12	.08	.12
7	-1.14	-1.19	.09	.10	.10	-1.18	.10	.06	.10	-1.15	.06	.04	.06
8	.19	.10	.06	.08	.10	.11	.01	.05	.08	.13	.03	.04	.06
9	.08	.01	.09	.08	.11	.06	.06	.05	.06	.07	.06	.04	.06
10	-1.27	-1.19	.05	.10	.09	-1.19	.07	.06	.10	-1.21	.04	.04	.07
11	-.16	-.14	.08	.10	.08	-.15	.07	.06	.07	-.15	.07	.04	.07
12	-.09	-.17	.09	.10	.12	-.12	.10	.07	.10	-.09	.06	.04	.06
13	.58	.49	.21	.13	.22	.51	.15	.08	.16	.53	.09	.06	.10
14	-.22	-.20	.20	.13	.20	-.21	.08	.08	.08	-.21	.04	.05	.04
15	.30	.35	.11	.14	.12	.32	.07	.09	.07	.30	.07	.06	.07
total	-.20	-.25	.10	.11	.11	-.24	.07	.07	.08	-.24	.06	.06	.07

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E3b: i=15; δ_2, small *e*, large *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	-.10	-.09	.09	.11	.09	-.09	.02	.07	.02	-.09	.01	.05	.01
2	-.56	-.55	.07	.12	.07	-.55	.05	.08	.05	-.55	.04	.05	.04
3	.69	.65	.09	.11	.09	.66	.06	.08	.06	.67	.04	.05	.04
4	-.14	-.19	.16	.11	.16	-.14	.07	.07	.07	-.14	.07	.05	.07
5	-1.17	-1.27	.11	.11	.14	-1.20	.08	.07	.08	-1.18	.05	.05	.05
6	.01	.02	.15	.19	.15	.02	.05	.14	.05	.02	.06	.09	.06
7	-1.14	-1.17	.11	.10	.11	-1.16	.08	.06	.08	-1.15	.05	.04	.05
8	.19	.10	.04	.08	.09	.11	.01	.05	.08	.13	.03	.04	.06
9	.08	.08	.09	.08	.09	.08	.06	.05	.06	.08	.06	.04	.06
10	-1.27	-1.19	.06	.09	.10	-1.19	.05	.06	.09	-1.21	.03	.04	.06
11	-.16	-.14	.08	.10	.08	-.15	.06	.07	.06	-.15	.06	.05	.06
12	-.09	-.17	.12	.10	.14	-.09	.12	.07	.12	-.10	.07	.05	.07
13	.58	.49	.17	.13	.19	.51	.12	.09	.13	.53	.07	.06	.08
14	-.22	-.20	.18	.12	.18	-.21	.05	.08	.05	-.21	.05	.05	.05
15	-.55	.31	.09	.14	.25	.30	.07	.09	.25	.30	.06	.06	.25
total	-.20	-.22	.11	.11	.11	-.24	.07	.08	.08	-.24	.04	.05	.05

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E3c: i=15; δ_2, large *e*, small *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	-.10	-.09	.07	.09	.07	-.09	.05	.06	.05	-.09	.05	.04	.05
2	-.56	-.55	.09	.11	.09	-.55	.04	.08	.04	-.55	.05	.05	.05
3	.69	.60	.13	.09	.15	.62	.07	.06	.09	.67	.06	.04	.06
4	-.14	-.17	.10	.08	.10	-.15	.05	.06	.05	-.14	.06	.04	.06
5	-1.17	-1.18	.08	.09	.08	-1.17	.06	.06	.06	-1.17	.03	.04	.03
6	.01	.02	.36	.19	.36	.02	.12	.14	.12	.02	.12	.07	.12
7	-1.14	-1.17	.09	.08	.09	-1.16	.07	.05	.07	-1.15	.04	.03	.04
8	.19	.10	.04	.06	.09	.11	.02	.05	.08	.13	.04	.03	.07
9	.08	.09	.05	.07	.05	.10	.04	.05	.04	.09	.04	.03	.04
10	-1.27	-1.19	.07	.08	.10	-1.19	.05	.05	.09	-1.21	.02	.03	.06
11	-.16	-.14	.06	.08	.06	-.15	.05	.06	.05	-.15	.06	.03	.06
12	-.09	-.15	.11	.08	.12	-.11	.05	.06	.05	-.10	.05	.03	.05
13	.58	.49	.13	.11	.15	.51	.09	.07	.11	.53	.06	.05	.07
14	-.22	-.20	.13	.10	.13	-.21	.05	.07	.05	-.21	.03	.04	.03
15	.30	.35	.10	.12	.11	.31	.04	.08	.04	.31	.03	.05	.03
total	-.20	-.25	.12	.10	.13	-.24	.08	.07	.08	-.24	.06	.04	.07

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E3d: i=15; δ_2, large *e*, large *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	-.10	-.07	.05	.09	.05	-.09	.04	.06	.04	-.09	.05	.04	.05
2	-.56	-.55	.07	.11	.07	-.55	.06	.07	.06	-.56	.04	.05	.04
3	.69	.65	.06	.10	.07	.66	.02	.06	.03	.67	.05	.04	.05
4	-.14	-.19	.10	.09	.11	-.15	.07	.06	.07	-.14	.07	.04	.07
5	-1.17	-1.18	.06	.07	.06	-1.17	.06	.06	.06	-1.17	.02	.04	.02
6	.01	.02	.09	.16	.09	.02	.10	.13	.10	.02	.10	.08	.10
7	-1.14	-1.17	.10	.06	.10	-1.16	.06	.05	.06	-1.15	.04	.03	.04
8	.19	.10	.05	.09	.10	.11	.05	.04	.09	.13	.03	.03	.06
9	.08	.12	.06	.09	.07	.10	.06	.04	.06	.08	.04	.03	.04
10	-1.27	-1.19	.04	.06	.08	-1.19	.04	.05	.08	-1.21	.03	.03	.06
11	-.16	-.14	.04	.09	.04	-.15	.05	.05	.05	-.15	.06	.04	.06
12	-.09	-.19	.08	.10	.12	-.17	.07	.05	.10	-.10	.05	.03	.05
13	.58	.49	.16	.12	.18	.51	.09	.07	.11	.53	.05	.05	.07
14	-.22	-.20	.13	.10	.13	-.21	.05	.06	.05	-.21	.05	.04	.05
15	.30	.33	.08	.17	.08	.32	.05	.07	.05	.31	.04	.05	.04
total	-.20	-.25	.09	.09	.10	-.24	.06	.06	.07	-.24	.05	.04	.06

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E3e: i=30; δ_2, small *e*, small *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	-.10	-.09	.09	.10	.09	-.09	.09	.07	.09	-.10	.06	.04	.06
2	-.56	-.55	.11	.12	.11	-.55	.02	.08	.02	-.56	.03	.05	.03
3	.69	.65	.13	.13	.13	.66	.13	.08	.13	.68	.11	.05	.11
4	-.14	-.19	.12	.11	.13	-.17	.06	.07	.06	-.16	.04	.05	.04
5	-1.17	-1.27	.06	.12	.11	-1.20	.04	.07	.05	-1.15	.03	.05	.03
6	.01	.02	.21	.18	.21	.02	.14	.12	.14	.01	.09	.08	.09
7	-1.14	-1.17	.07	.09	.07	-1.16	.06	.06	.06	-1.15	.03	.04	.03
8	.19	.10	.05	.08	.10	.15	.04	.05	.05	.17	.05	.03	.05
9	.08	.12	.12	.08	.12	.10	.04	.06	.04	.08	.03	.04	.03
10	-1.27	-1.19	.05	.10	.09	-1.19	.06	.06	.10	-1.22	.03	.04	.05
11	-.37	-.34	.18	.13	.18	-.34	.11	.09	.11	-.36	.06	.06	.06
12	-.20	-.19	.11	.12	.11	-.19	.04	.08	.04	-.20	.05	.05	.05
13	.26	.28	.15	.14	.15	.27	.09	.09	.09	.27	.05	.06	.05
14	-.32	-.30	.15	.14	.15	-.31	.10	.09	.10	-.31	.09	.06	.09
15	-.28	-.31	.04	.14	.05	-.29	.06	.09	.06	-.28	.04	.06	.04
16	-.16	-.16	.17	.10	.17	-.16	.07	.07	.07	-.16	.05	.04	.05
17	-.09	-.19	.10	.10	.14	-.17	.07	.07	.10	-.11	.06	.04	.06
18	.58	.59	.14	.13	.14	.59	.04	.09	.04	.58	.04	.06	.04
19	-.22	-.25	.19	.12	.19	-.23	.06	.08	.06	-.23	.07	.05	.07
20	.30	.35	.08	.14	.09	.34	.10	.09	.10	.31	.07	.06	.07
21	.94	1.01	.30	.17	.30	.99	.07	.11	.08	.98	.07	.07	.08
22	-.74	-.77	.13	.14	.13	-.75	.06	.09	.06	-.74	.03	.06	.03
23	.08	.05	.23	.20	.23	.07	.08	.14	.08	.07	.07	.09	.07

24	-.80	-.75	.17	.17	.17	-.76	.20	.12	.20	-.79	.09	.08	.07
25	-.88	-.92	.18	.18	.18	-.91	.10	.12	.10	-.89	.09	.08	.09
26	.49	.40	.27	.19	.28	.41	.11	.11	.13	.42	.07	.08	.09
27	-.90	-.99	.04	.18	.09	-.95	.06	.12	.07	-.94	.08	.08	.08
28	.93	1.01	.28	.15	.29	.99	.09	.09	.10	.96	.06	.07	.06
29	-.67	-.66	.17	.13	.17	-.66	.06	.09	.06	-.67	.03	.06	.03
30	-.38	-.42	.14	.13	.14	-.41	.06	.09	.06	-.39	.05	.06	.05
Total	-.20	-.15	.11	.13	.12	-.15	.09	.09	.10	-.17	.07	.06	.07

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E3f: i=30; δ_2, small *e*, large *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	-.10	-.09	.10	.10	.10	-.09	.10	.07	.07	-.10	.05	.05	.05
2	-.56	-.55	.04	.12	.04	-.55	.02	.08	.02	-.55	.02	.05	.02
3	.69	.66	.20	.13	.20	.66	.11	.08	.11	.68	.06	.05	.04
4	-.14	-.19	.07	.11	.08	-.17	.06	.07	.06	-.15	.02	.05	.02
5	-1.17	-1.27	.10	.12	.14	-1.22	.04	.08	.06	-1.20	.04	.05	.05
6	.01	.05	.26	.18	.26	.05	.22	.14	.22	.02	.13	.09	.13
7	-1.14	-1.20	.09	.09	.10	-1.16	.07	.06	.07	-1.15	.04	.04	.04
8	.19	.10	.05	.08	.10	.15	.06	.06	.07	.17	.05	.04	.05
9	.08	.09	.10	.08	.10	.09	.07	.06	.07	.08	.04	.04	.04
10	-1.27	-1.19	.14	.10	.16	-1.19	.05	.06	.09	-1.21	.01	.04	.06
11	-.37	-.34	.06	.13	.06	-35	.11	.09	.11	-.34	.02	.06	.03
12	-.20	-.18	.08	.12	.08	-.19	.07	.08	.07	-.20	.03	.05	.03
13	.26	.29	.14	.14	.14	.28	.12	.09	.12	.27	.03	.06	.03
14	-.32	-.30	.16	.14	.16	-.31	.11	.09	.11	-.31	.07	.06	.07
15	-.28	-.31	.09	.14	.09	-.29	.06	.09	.06	-.28	.06	.06	.06
16	-.16	-.17	.11	.10	.11	-.16	.05	.07	.05	-.16	.05	.05	.05
17	-.09	-.11	.13	.10	.13	-.10	.07	.07	.07	-.09	.06	.04	.06
18	.58	.59	.13	.13	.13	.59	.10	.09	.10	.58	.04	.06	.04
19	-.22	-.25	.12	.12	.12	-.25	.07	.08	.07	-.23	.04	.06	.04
20	.30	.37	.09	.14	.11	.32	.09	.09	.09	.31	.05	.06	.05
21	.94	1.01	.10	.17	.12	.99	.05	.12	.07	.95	.06	.08	.06
22	-.74	-.76	.18	.14	.18	-.75	.06	.10	.06	-.74	.02	.07	.02
23	.08	.06	.09	.20	.09	.07	.11	.14	.11	.08	.15	.09	.15

24	-.80	-.75	.14	.17	.14	-.75	.16	.12	.16	-.76	.11	.08	.11
25	-.88	-.92	.18	.18	.18	-.90	.15	.12	.15	-.89	.10	.08	.10
26	.49	.48	.13	.19	.13	.48	.11	.12	.11	.49	.03	.08	.03
27	-.90	-.93	.14	.18	.14	-.92	.06	.12	.06	-.91	.04	.09	.04
28	.93	.94	.15	.15	.15	.93	.11	.10	.11	.93	.08	.07	.08
29	-.67	-.66	.12	.13	.12	-.66	.05	.09	.05	-.66	.04	.06	.04
30	-.38	-.40	.03	.13	.03	-.40	.06	.09	.06	-.39	.02	.06	.02
total	-.20	-.14	.10	.14	.11	-.16	.08	.09	.08	-.17	.05	.06	.05

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E3g: i=30; δ_2, large *e*, small *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	-.10	-.09	.09	.08	.09	-.09	.07	.06	.07	-.10	.03	.04	.03
2	-.56	-.53	.06	.12	.06	-.55	.04	.08	.04	-.55	.04	.05	.04
3	.69	.65	.13	.10	.13	.66	.09	.07	.09	.67	.06	.04	.06
4	-.14	-.24	.10	.09	.14	-.22	.06	.06	.10	-.16	.04	.04	.04
5	-1.17	-1.27	.09	.09	.13	-1.22	.04	.06	.06	-1.20	.04	.04	.05
6	.01	.03	.18	.19	.18	.03	.15	.14	.15	.02	.10	.08	.10
7	-1.14	-1.17	.10	.07	.10	-1.16	.03	.05	.03	-1.15	.03	.03	.03
8	.19	.10	.05	.06	.10	.15	.07	.05	.08	.17	.04	.03	.04
9	.08	.15	.07	.06	.09	.11	.06	.05	.06	.09	.03	.03	.03
10	-1.27	-1.19	.11	.08	.13	-1.19	.03	.06	.08	-1.21	.02	.03	.06
11	-.37	-.34	.09	.11	.09	-.35	.13	.07	.13	-.36	.05	.05	.05
12	-.20	-.17	.07	.09	.07	-.19	.06	.06	.06	-.20	.02	.04	.02
13	.26	.28	.05	.11	.05	.27	.09	.08	.09	.26	.04	.05	.04
14	-.32	-.28	.08	.11	.08	-.29	.03	.07	.04	-.31	.04	.05	.04
15	-.28	-.31	.06	.10	.06	-.30	.04	.07	.04	-.28	.03	.05	.03
16	-.16	-.20	.09	.08	.09	-.19	.06	.06	.06	-.18	.04	.04	.04
17	-.09	-.19	.12	.07	.15	-.17	.06	.06	.10	-.12	.05	.03	.05
18	.58	.60	.11	.10	.11	.59	.06	.07	.06	.58	.02	.05	.02
19	-.22	-.23	.11	.10	.11	-.23	.04	.07	.04	-.22	.05	.04	.05
20	.30	.39	.11	.13	.14	.37	.07	.08	.09	.37	.04	.05	.08
21	.94	1.01	.09	.13	.11	.99	.08	.08	.09	.97	.08	.06	.08
22	-.74	-.82	.12	.10	.14	-.80	.06	.07	.08	-.79	.01	.05	.05
23	.08	.05	.11	.17	.11	.07	.10	.12	.10	.11	.10	.08	.10

24	-.80	-.75	.08	.13	.09	-.76	.08	.09	.08	-.76	.06	.06	.07
25	-.88	-.98	.10	.15	.14	-.95	.09	.10	.11	-.95	.09	.06	.11
26	.49	.46	.18	.16	.18	.46	.12	.10	.12	.48	.09	.07	.09
27	-.90	-.99	.12	.14	.15	-.97	.05	.10	.08	-.95	.04	.07	.06
28	.93	.94	.15	.11	.15	.93	.12	.08	.12	.93	.06	.05	.06
29	-.67	-.66	.09	.10	.09	-.66	.06	.07	.06	-.67	.04	.04	.04
30	-.38	-.42	.07	.09	.08	-.41	.09	.07	.09	-.39	.04	.04	.04
total	-.20	-.14	.09	.11	.10	-.15	.07	.08	.08	-.16	.05	.05	.06

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E3h: i=30; δ_2, large *e*, large *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	-.10	-.08	.12	.08	.12	-.09	.11	.06	.11	-.09	.06	.04	.06
2	-.56	-.57	.06	.11	.06	-.57	.06	.07	.06	-.55	.06	.05	.06
3	.69	.65	.13	.10	.13	.66	.09	.06	.09	.67	.07	.04	.07
4	-.14	-.24	.09	.09	.13	-.22	.08	.06	.11	-.19	.02	.04	.05
5	-1.17	-1.20	.08	.09	.10	-1.19	.05	.06	.05	-1.18	.03	.04	.03
6	.01	.02	.15	.20	.15	.02	.15	.11	.15	.01	.11	.08	.11
7	-1.14	-1.25	.07	.07	.13	-1.22	.04	.05	.08	-1.20	.03	.03	.06
8	.19	.10	.05	.06	.10	.15	.04	.04	.05	.16	.04	.03	.05
9	.08	.14	.07	.07	.09	.13	.07	.04	.08	.13	.03	.03	.05
10	-1.27	-1.18	.14	.08	.16	-1.19	.06	.05	.10	-1.21	.03	.03	.06
11	-.37	-.34	.09	.11	.09	-.35	.13	.08	.13	-.35	.05	.05	.05
12	-.20	-.17	.04	.09	.05	-.18	.07	.06	.07	-.19	.03	.04	.03
13	.26	.28	.08	.11	.08	.27	.10	.07	.10	.27	.07	.05	.07
14	-.32	-.30	.08	.11	.08	-.31	.07	.07	.07	-.31	.05	.05	.05
15	-.28	-.31	.10	.11	.10	-.29	.07	.07	.07	-.28	.04	.04	.04
16	-.16	-.26	.08	.08	.12	-.24	.06	.05	.10	-.24	.04	.04	.08
17	-.09	-.18	.10	.07	.13	-.16	.05	.05	.08	-.12	.04	.03	.05
18	.58	.62	.16	.11	.16	.60	.06	.07	.06	.59	.02	.05	.02
19	-.22	-.25	.03	.10	.04	-.24	.03	.07	.03	-.22	.03	.04	.03
20	.30	.37	.07	.12	.09	.36	.09	.09	.10	.36	.05	.05	.07
21	.94	1.01	.06	.13	.09	.99	.07	.09	.08	.95	.06	.06	.06
22	-.74	-.85	.09	.10	.14	-.81	.05	.07	.08	-.79	.03	.05	.05
23	.08	.05	.06	.17	.06	.07	.14	.12	.14	.08	.12	.08	.12

24	-.80	-.75	.09	.13	.10	-.75	.10	.09	.11	-.76	.07	.06	.08
25	-.88	-.91	.09	.14	.09	-.89	.10	.09	.10	-.88	.10	.06	.10
26	.49	.52	.20	.15	.20	.51	.15	.09	.15	.50	.09	.07	.09
27	-.90	-.97	.17	.15	.18	-.96	.07	.09	.09	-.93	.06	.07	.06
28	.93	.95	.15	.11	.15	.95	.15	.08	.15	.93	.05	.05	.05
29	-.67	-.66	.08	.09	.08	-.66	.05	.06	.05	-.66	.02	.04	.02
30	-.38	-.46	.05	.09	.09	-.42	.07	.07	.08	-.41	.03	.04	.04
total	-.20	-.13	.11	.11	.13	-.17	.09	.07	.09	-.18	.06	.05	.06

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E4a: i=15; δ_3, small *e*, small *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	.45	.46	.10	.10	.10	.46	.06	.07	.06	.46	.05	.05	.05
2	-.52	-.56	.13	.12	.13	-.52	.09	.09	.09	-.53	.07	.06	.07
3	1.05	1.14	.13	.12	.15	1.04	.11	.08	.11	1.04	.06	.06	.06
4	.57	.55	.16	.11	.16	.54	.10	.07	.10	.55	.05	.05	.05
5	-.04	-.14	.05	.09	.11	-.08	.09	.06	.09	-.05	.06	.04	.06
6	.40	.39	.15	.17	.15	.39	.11	.11	.11	.38	.05	.08	.05
7	-.02	-.06	.03	.08	.05	-.03	.05	.05	.05	-.03	.03	.03	.03
8	1.24	1.31	.17	.11	.18	1.28	.09	.07	.09	1.24	.07	.05	.07
9	1.02	1.04	.06	.10	.06	1.05	.07	.06	.07	1.03	.04	.04	.04
10	-.08	-.12	.07	.08	.08	-.10	.06	.05	.06	-.08	.04	.04	.04
11	.87	.94	.07	.11	.09	.89	.08	.07	.08	.90	.05	.05	.05
12	.97	1.07	.08	.12	.12	1.06	.03	.08	.09	1.02	.02	.05	.05
13	1.34	1.41	.20	.14	.21	1.41	.09	.09	.11	1.33	.08	.06	.08
14	.32	.24	.11	.13	.13	.34	.06	.08	.06	.33	.03	.06	.03
15	1.95	2.12	.33	.18	.37	1.99	.18	.12	.18	1.92	.12	.08	.12
total	.63	.65	.11	.12	.11	.65	.08	.08	.08	.63	.06	.05	.06

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E4b: i=15; δ_3, small *e*, large *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	.45	.43	.10	.11	.10	.44	.07	.07	.07	.43	.06	.05	.06
2	-.52	-.59	.15	.13	.16	-.55	.08	.09	.08	-.53	.06	.06	.06
3	1.05	1.05	.08	.13	.08	1.02	.08	.09	.08	1.04	.05	.06	.05
4	.57	.54	.18	.11	.18	.54	.11	.07	.11	.56	.06	.05	.06
5	-.04	-.08	.06	.09	.07	-.05	.07	.06	.07	-.03	.04	.04	.04
6	.40	.31	.11	.17	.14	.32	.03	.12	.08	.37	.05	.08	.05
7	-.02	-.04	.04	.08	.04	-.03	.05	.05	.05	-.02	.03	.03	.03
8	1.24	1.33	.12	.11	.15	1.28	.06	.07	.07	1.24	.05	.05	.05
9	1.02	1.05	.06	.10	.06	1.05	.05	.07	.05	1.04	.04	.05	.04
10	-.08	-.11	.05	.08	.05	-.10	.05	.05	.05	-.08	.03	.04	.03
11	.87	.92	.07	.11	.08	.91	.08	.08	.08	.91	.06	.05	.07
12	.97	1.07	.04	.12	.10	1.05	.03	.08	.08	1.02	.02	.06	.05
13	1.34	1.48	.18	.15	.22	1.43	.08	.10	.12	1.34	.08	.07	.08
14	.32	.33	.14	.12	.14	.33	.07	.08	.07	.32	.05	.06	.05
15	1.95	2.13	.31	.19	.35	1.99	.20	.12	.20	1.92	.11	.08	.11
total	.63	.65	.11	.12	.11	.64	.09	.08	.09	.63	.06	.06	.06

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E4c: i=15; δ_3, large *e*, small *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	.45	.45	.07	.09	.07	.47	.01	.06	.01	.45	.01	.04	.01
2	-.52	-.59	.15	.12	.16	-.52	.09	.08	.09	-.55	.06	.05	.06
3	1.05	1.06	.10	.10	.10	1.06	.09	.07	.09	1.04	.03	.05	.03
4	.57	.60	.13	.09	.13	.57	.08	.06	.08	.57	.05	.04	.05
5	-.04	-.06	.06	.07	.06	-.03	.08	.05	.08	-.03	.05	.03	.05
6	.40	.39	.19	.18	.19	.38	.06	.13	.06	.41	.04	.07	.04
7	-.02	-.03	.04	.07	.04	-.01	.05	.05	.05	-.02	.03	.03	.03
8	1.24	1.32	.12	.09	.14	1.28	.05	.06	.06	1.24	.03	.04	.03
9	1.02	1.08	.05	.08	.07	1.06	.05	.06	.06	1.04	.04	.04	.04
10	-.08	-.12	.04	.07	.05	-.09	.03	.05	.03	-.08	.02	.03	.02
11	.87	.96	.08	.09	.12	.92	.07	.06	.08	.90	.06	.04	.06
12	.97	1.06	.09	.09	.12	1.06	.04	.07	.09	1.02	.04	.04	.12
13	1.34	1.50	.15	.12	.21	1.41	.09	.08	.11	1.35	.06	.05	.06
14	.32	.35	.12	.10	.12	.35	.07	.07	.07	.35	.03	.04	.04
15	1.95	2.12	.28	.16	.32	2.01	.15	.10	.16	1.93	.08	.07	.08
total	.63	.67	.11	.10	.11	.66	.07	.07	.07	.64	.04	.04	.04

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E4d: i=15; δ_3, large *e*, large *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	.45	.47	.09	.09	.11	.46	.06	.06	.06	.44	.02	.04	.02
2	-.52	-.54	.07	.11	.07	-.54	.07	.07	.07	-.54	.05	.05	.05
3	1.05	1.10	.12	.10	.13	1.06	.08	.07	.08	1.05	.04	.05	.04
4	.57	.61	.11	.09	.11	.58	.09	.06	.09	.56	.04	.04	.04
5	-.04	-.05	.04	.07	.04	-.03	.06	.05	.06	-.04	.04	.03	.04
6	.40	.44	.12	.16	.12	.38	.09	.12	.09	.40	.04	.08	.04
7	-.02	-.02	.04	.06	.04	-.02	.05	.04	.05	-.03	.03	.03	.03
8	1.24	1.33	.10	.09	.13	1.29	.05	.06	.07	1.24	.03	.04	.03
9	1.02	1.08	.09	.09	.10	1.05	.06	.06	.06	1.03	.05	.04	.05
10	-.08	-.11	.04	.06	.05	-.10	.04	.04	.04	-.09	.03	.03	.03
11	.87	.95	.07	.09	.10	.90	.07	.06	.07	.89	.06	.04	.06
12	.97	1.05	.07	.10	.10	1.04	.06	.07	.09	1.01	.04	.04	.05
13	1.34	1.46	.16	.12	.20	1.40	.09	.08	.10	1.34	.07	.05	.07
14	.32	.35	.10	.10	.10	.35	.05	.07	.05	.33	.04	.04	.04
15	1.95	2.15	.20	.17	.28	2.01	.13	.11	.14	1.93	.09	.07	.09
total	.63	.68	.11	.10	.12	.65	.08	.07	.08	.63	.05	.05	.05

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E4e: i=30; δ_3, small *e*, small *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	.45	.48	.09	.10	.09	.47	.02	.07	.02	.45	.04	.04	.04
2	-.52	-.59	.11	.13	.13	-.54	.05	.08	.05	-.50	.04	.06	.04
3	1.05	1.04	.13	.14	.13	1.01	.13	.09	.13	1.02	.08	.06	.08
4	.57	.55	.12	.12	.12	.53	.07	.08	.08	.55	.05	.05	.05
5	-.04	-.02	.06	.09	.06	-.05	.05	.06	.05	-.05	.02	.04	.02
6	.40	.33	.21	.17	.22	.28	.05	.11	.13	.41	.02	.07	.02
7	-.02	.00	.07	.07	.07	-.01	.04	.05	.04	-.02	.02	.03	.02
8	1.24	1.32	.05	.10	.09	1.26	.07	.07	.07	1.25	.06	.05	.06
9	1.02	1.05	.12	.10	.12	1.00	.10	.07	.10	1.01	.06	.04	.06
10	-.08	-.14	.05	.08	.07	-.11	.02	.06	.03	-.10	.02	.04	.02
11	.25	.36	.18	.13	.21	.33	.09	.09	.12	.26	.03	.06	.03
12	.53	.50	.11	.12	.11	.52	.06	.08	.06	.53	.04	.05	.04
13	.92	1.01	.15	.15	.17	1.00	.12	.10	.14	.92	.07	.07	.07
14	.73	.79	.15	.14	.16	.72	.12	.09	.12	.74	.10	.06	.10
15	.62	.67	.04	.11	.06	.64	.06	.08	.06	.64	.05	.05	.05
16	.87	.89	.17	.12	.17	.89	.07	.08	.07	.86	.06	.05	.06
17	.97	1.09	.10	.12	.15	1.04	.06	.08	.09	1.00	.04	.05	.05
18	1.34	1.41	.14	.13	.15	1.38	.05	.09	.06	1.35	.06	.06	.06
19	.32	.31	.19	.13	.19	.29	.12	.09	.12	.28	.11	.06	.11
20	1.92	2.14	.08	.18	.23	2.01	.05	.12	.10	1.93	.05	.08	.05
21	1.52	1.72	.30	.21	.36	1.62	.13	.14	.17	1.60	.10	.10	.12
22	.10	.09	.13	.12	.13	.04	.11	.08	.11	.08	.05	.06	.05
23	.36	.30	.23	.20	.23	.28	.17	.13	.18	.33	.06	.09	.06

24	-.20	-.22	.17	.14	.17	-.21	.10	.09	.10	-.20	.08	.06	.08
25	-.70	-.80	.18	.19	.20	-.76	.20	.13	.20	-.71	.10	.09	.10
26	1.05	1.15	.27	.19	.28	1.12	.15	.12	.16	1.07	.08	.08	.08
27	.12	.06	.04	.11	.07	.07	.07	.07	.08	.09	.07	.05	.07
28	1.61	1.74	.28	.20	.30	1.70	.21	.13	.22	1.63	.13	.09	.13
29	.12	.10	.17	.12	.17	.10	.05	.08	.05	.12	.03	.05	.03
30	.27	.24	.14	.13	.14	.26	.13	.09	.13	.24	.08	.06	.08
Total	.55	.59	.14	.13	.14	.56	.08	.09	.08	.56	.06	.06	.06

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E4f: i=30; δ_3, small *e*, large *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	.45	.44	.11	.11	.11	.43	.04	.07	.04	.43	.05	.05	.05
2	-.52	-.63	.13	.13	.17	-.58	.06	.08	.07	-.52	.05	.06	.05
3	1.05	1.01	.11	.14	.11	1.00	.09	.10	.10	1.02	.05	.06	.05
4	.57	.57	.16	.12	.16	.54	.09	.08	.09	.53	.02	.05	.04
5	-.04	-.03	.07	.09	.07	-.06	.05	.06	.05	-.04	.03	.04	.03
6	.40	.36	.19	.17	.19	.30	.11	.12	.14	.34	.04	.09	.08
7	-.02	.00	.10	.08	.10	-.02	.05	.05	.05	-.03	.02	.03	.02
8	1.24	1.32	.06	.11	.10	1.26	.07	.07	.07	1.27	.05	.05	.05
9	1.02	1.03	.11	.10	.11	.98	.09	.07	.09	.97	.03	.04	.05
10	-.08	-.13	.06	.08	.07	-.11	.03	.06	.04	-.11	.01	.04	.03
11	.25	.37	.19	.14	.22	.32	.10	.10	.12	.25	.03	.06	.03
12	.53	.51	.08	.12	.08	.51	.05	.08	.05	.55	.04	.06	.04
13	.92	1.01	.13	.16	.15	.97	.12	.10	.13	.95	.03	.07	.04
14	.73	.76	.10	.15	.10	.71	.13	.09	.13	.63	.04	.06	.10
15	.62	.69	.05	.12	.08	.65	.07	.08	.07	.63	.05	.05	.05
16	.87	.89	.13	.12	.13	.87	.06	.08	.06	.83	.03	.06	.05
17	.97	1.09	.09	.12	.15	1.03	.07	.08	.09	.98	.04	.05	.04
18	1.34	1.40	.13	.14	.14	1.36	.06	.09	.06	1.37	.06	.07	.06
19	.32	.32	.12	.12	.12	.30	.08	.08	.08	.29	.09	.06	.09
20	1.92	2.13	.09	.18	.22	2.01	.05	.12	.10	1.95	.03	.09	.04
21	1.52	1.65	.28	.22	.30	1.57	.12	.15	.13	1.53	.03	.10	.03
22	.10	.08	.10	.13	.10	.02	.09	.09	.12	.05	.03	.06	.05
23	.36	.34	.18	.19	.18	.27	.13	.14	.15	.38	.06	.09	.04

24	-.20	-.21	.15	.13	.15	-.20	.08	.09	.08	-.23	.01	.06	.03
25	-.70	-.80	.15	.20	.18	-.75	.13	.13	.13	-.70	.07	.08	.07
26	1.05	1.11	.22	.20	.22	1.08	.15	.13	.15	1.05	.07	.08	.07
27	.12	.06	.04	.11	.07	.08	.05	.07	.06	.06	.05	.05	.07
28	1.61	1.66	.27	.20	.27	1.66	.20	.14	.20	1.5	.10	.09	.14
29	.12	.11	.16	.12	.16	.10	.05	.08	.05	.10	.01	.05	.02
30	.27	.28	.06	.13	.06	.30	.13	.09	.13	.26	.08	.06	.08
total	.55	.58	.13	.14	.13	.55	.09	.09	.09	.54	.04	.06	.05

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E4g: i=30; δ_3, large *e*, small *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	.45	.44	.03	.09	.03	.43	.03	.06	.03	.44	.01	.04	.01
2	-.52	-.67	.13	.12	.19	-.59	.11	.08	.13	-.55	.03	.05	.04
3	1.05	1.07	.10	.11	.10	1.07	.06	.08	.06	1.04	.05	.05	.05
4	.57	.52	.18	.09	.18	.57	.07	.06	.07	.54	.03	.04	.04
5	-.04	-.03	.08	.07	.08	-.02	.06	.05	.06	-.05	.03	.03	.03
6	.40	.31	.20	.18	.21	.32	.04	.12	.08	.36	.07	.07	.08
7	-.02	-.00	.08	.06	.08	.00	.05	.05	.05	-.02	.01	.03	.01
8	1.24	1.30	.04	.09	.07	1.25	.08	.06	.08	1.25	.05	.04	.05
9	1.02	1.06	.09	.08	.09	1.04	.04	.06	.04	1.02	.03	.04	.03
10	-.08	-.11	.03	.06	.04	-.09	.01	.05	.01	-.09	.02	.03	.02
11	.25	.35	.12	.11	.15	.32	.05	.08	.08	.25	.02	.05	.02
12	.53	.53	.11	.09	.11	.58	.06	.06	.07	.55	.04	.04	.04
13	.92	1.01	.10	.12	.13	.98	.05	.08	.07	.93	.07	.05	.07
14	.73	.77	.19	.11	.19	.76	.11	.07	.11	.74	.07	.05	.07
15	.62	.67	.04	.09	.06	.65	.06	.06	.06	.64	.03	.04	.03
16	.87	.87	.13	.09	.13	.88	.09	.06	.09	.86	.04	.04	.04
17	.97	1.08	.07	.09	.13	1.05	.05	.07	.09	1.00	.04	.04	.05
18	1.34	1.45	.11	.11	.15	1.42	.06	.08	.10	1.37	.06	.05	.06
19	.32	.32	.13	.10	.13	.30	.11	.07	.11	.29	.06	.04	.06
20	1.92	2.18	.09	.17	.27	2.04	.03	.11	.12	1.93	.04	.07	.04
21	1.52	1.64	.24	.16	.26	1.60	.06	.11	.10	1.55	.06	.08	.06
22	.10	.05	.09	.09	.10	.06	.08	.07	.08	.08	.03	.04	.03
23	.36	.32	.13	.16	.13	.25	.08	.12	.13	.33	.06	.07	.06

24	-.20	-.20	.13	.10	.13	-.18	.06	.07	.06	-.20	.03	.04	.03
25	-.70	-.93	.21	.17	.31	-.76	.18	.11	.18	-.75	.07	.07	.08
26	1.05	1.06	.18	.16	.18	1.11	.10	.11	.11	1.05	.06	.07	.06
27	.12	.06	.05	.08	.07	.09	.04	.06	.05	.09	.03	.03	.04
28	1.61	1.69	.28	.15	.28	1.74	.18	.11	.22	1.64	.11	.07	.11
29	.12	.07	.17	.09	.17	.12	.04	.06	.04	.11	.02	.04	.02
30	.27	.28	.08	.10	.08	.29	.05	.07	.05	.26	.04	.05	.04
total	.55	.57	.12	.11	.12	.57	.08	.08	.08	.55	.05	.06	.05

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

Table E4h: i=30; δ_3, large *e*, large *f*

		N=500 Estimates				N=1000 Estimates				N=2000 Estimates			
Item	True	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE	Mean	SD	SE	RMSE
1	.45	.47	.07	.08	.07	.46	.06	.06	.06	.44	.04	.04	.04
2	-.52	-.63	.11	.12	.15	-.58	.05	.08	.07	-.54	.01	.05	.02
3	1.05	1.10	.10	.11	.11	1.09	.04	.08	.05	1.04	.06	.05	.06
4	.57	.51	.15	.10	.16	.52	.10	.06	.11	.54	.04	.04	.05
5	-.04	-.02	.03	.07	.03	-.04	.05	.05	.05	-.06	.03	.03	.03
6	.40	.33	.16	.18	.17	.36	.11	.11	.11	.41	.03	.08	.03
7	-.02	-.00	.07	.06	.07	-.01	.05	.04	.05	-.03	.01	.03	.01
8	1.24	1.32	.05	.09	.09	1.26	.08	.06	.08	1.25	.05	.04	.05
9	1.02	1.05	.08	.08	.08	1.02	.09	.05	.09	1.02	.03	.04	.03
10	-.08	-.12	.02	.06	.04	-.09	.03	.04	.03	-.10	.01	.03	.02
11	.25	.36	.13	.11	.17	.31	.05	.07	.07	.26	.02	.05	.02
12	.53	.54	.11	.09	.11	.56	.06	.06	.06	.55	.04	.04	.04
13	.92	1.03	.08	.12	.13	1.01	.04	.08	.09	.93	.05	.06	.05
14	.73	.78	.16	.11	.16	.75	.10	.07	.10	.73	.07	.05	.07
15	.62	.68	.02	.09	.06	.64	.05	.06	.05	.64	.04	.04	.04
16	.87	.90	.12	.09	.12	.89	.08	.06	.08	.87	.05	.04	.05
17	.97	1.09	.07	.09	.13	1.05	.05	.06	.09	1.01	.05	.04	.06
18	1.34	1.41	.07	.11	.09	1.39	.06	.08	.07	1.35	.05	.05	.05
19	.32	.30	.15	.10	.15	.32	.13	.07	.13	.29	.08	.05	.08
20	1.92	2.17	.12	.16	.27	2.02	.06	.11	.11	1.93	.05	.07	.05
21	1.52	1.68	.26	.18	.30	1.51	.12	.11	.12	1.52	.04	.08	.04
22	.10	.05	.09	.09	.10	.04	.08	.06	.10	.06	.03	.04	.05
23	.36	.33	.08	.17	.08	.31	.07	.11	.08	.32	.06	.07	.06

24	-.20	-.18	.10	.10	.10	-.17	.09	.07	.09	-.20	.05	.05	.05
25	-.70	-.86	.21	.15	.26	-.79	.12	.10	.15	-.73	.07	.07	.07
26	1.05	1.11	.18	.16	.18	1.14	.11	.10	.14	1.06	.08	.07	.08
27	.12	.07	.05	.08	.07	.08	.04	.05	.05	.09	.04	.03	.05
28	1.61	1.67	.22	.16	.22	1.66	.18	.11	.18	1.61	.11	.07	.11
29	.12	.10	.15	.09	.15	.09	.05	.06	.05	.10	.03	.04	.03
30	.27	.28	.03	.10	.03	.29	.07	.07	.07	.28	.03	.04	.03
total	.55	.58	.11	.11	.11	.57	.08	.07	.08	.55	.06	.05	.06

Note: SD = Variation of iteration values within a chain, across simulations

Note: SE = Variation of iteration values in a chain (Posterior Standard Deviation; MCMC SE)

Note: RMSE = Root of the squared deviations of the estimated iteration values around the true value

APPENDIX F: Sample items

NEUROTICISM

1 I often feel jittery and tense

2 I am often nervous and tense

3 When I am under great stress, I often feel like I am about to break down

4 I am always worried about how things might go wrong

5 I am often sad and depressed

AGREEABLENESS

1 I try to be kind to everyone I know

2 I always treat other people with kindness

3 I am always considerate of the feeling of others

4 I am considered by others to be a very friendly person

5 I try to be pleasant in every situation

CONSCIENTIOUSNESS

1 I like to keep all my belongings neat and organized

2 I like to have a place for everything and everything in its place

3 I try to set a schedule for accomplishing tasks, and stick to it

4 If I start something, I work until it is finished to my satisfaction

5 If I commit myself to do something, I always carry through

EXTROVERSION

1 I am a very shy person

2 At social functions, I talk to as many people as possible

3 Most of my friends would describe me as a "talker"

4 My friends consider me to be bashful

5 If things get too boring at a party, I try to get things going

OPENNESS

1 I spend a lot of time in meditation and deep thought

2 Philosophical discussions bore me

3 I would enjoy being a theoretical scientist

4 I have thought a lot about the origin of the universe

5 I have a lot of intellectual curiosity

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19,* 6, 716-723.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43,* 561-573.

Bargh, J. A. (1982). Attention and automaticity in the processing of self-relevant information. *Journal of Personality and Social Psychology, 43,* 425-436.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scores in two or more nominal categories. *Psychometrika, 37,* 29-51.

Bock, R. D., Muraki, E., & Pfeiffenberger, W. (1988). Item pool maintenance in the presence of item parameter drift. *Journal of Educational Measurement, 25,* 275-285.

Cervone, D., Shadel, W. G., & Jencius, S. (2001). Social-cognitive theory of personality assessment. *Personality and Social Psychology Review, 5,* 33-51.

Chernyshenko, O. S., Stark, S., Chan, K-Y., Drasgow, F., & Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research, 36,* 4, 523-562.

Christal, R. E. (1993). R&D Summary report F33615-91-D-0010. Armstrong Laboratories, Brooks AFB.

Costa, P.T., Jr. & McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual.* Odessa, FL: Psychological Assessment Resources.

Cramer, D. (1993). Perceived and desired facilitativeness of one's closest friend, need for approval and self-esteem. *British Journal of Medical Psychology, 66,* 97-104.

Crocker, L. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Harcourt Brace.

Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Fekken, G. C., & Holden, R. R. (1992). Response latency evidence for viewing personality traits as schema indicators. *Journal of Research in Personality, 26,* 103-120.

Feldman, J. M., & Lynch, Jr., J. G. (1988). Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *Journal of Applied Psychology, 73,* 3, 421-435.

Ferrando, P. J., & Anguiano-Carrasco, C. (2009). Assessing the impact of faking on binary personality measures: An IRT-based multiple-group factor analytic procedure. *Multivariate Behavioral Research, 44,* 497-524.

Gelman, A. and Rubin, D. B. (1992). Inferences from iterative simulation using multiple sequences (with discussion). *Statistical Science, 7*, 4, 457-511.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. M. Bernardo, J. O. Berger, A. P. David, & A. F. M. Smith (Eds.), *Bayesian statistics* (Vol. 4, pp. 169-193). Oxford, UK: Oxford University Press.

Goldberg, L. R. (1978). The reliability of reliability: The generality and correlates of intra-individual consistency in responses to structured personality inventories. *Applied Psychological Measurement, 2*, 269-291.

Goldstein, H. (1983). Measuring changes in educational attainment over time: Problems and possibilities. *Journal of Educational Measurement, 20*, 4, 369-377.

Greenwald, A. G. & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*, 1, 4-27.

Haario, H., Saksman, E., and Tamminen, J. (2001), An adaptive Metropolis algorithm. *Bernoulli, 7*, 2, 223-242.

Hamilton, J. C., & Shuminsky, T. R. (1990). Self-awareness mediates the relationship between serial position and item reliability. *Journal of Personality and Social Psychology, 59*, 6, 1301-1307.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications.* Boston: Kluwer-Nijhoff.

Harris, J. A., and Lucia, A. (2003). The relationship between self-report model and personality, *Personality and Individual Differences, 35*, 8, 1903-1909.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57*, 97-109.

Hayes, D. P. (1964). Item order and Guttman scales. *American Journal of Sociology, 20*, 52-58.

Henry, M. S., & Raju, N. S. (2006). The effects of traited and situational impression management on a personality test: An empirical analysis. *Psychology Science, 48*, 247-267.

Holden, R. R., Kroner, D. G., Fekken, G. C., & Popham, S. M. (1992). A model of personality test item response dissimulation. *Journal of Personality and Social Psychology, 63*, 272-279.

Holtgraves, T. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin, 30*, 2, 161-172.

Kim, Jee-Seon, & Bolt, D. M. (2007). Estimating item response theory models using Markov Chain Monte Carlo methods. *Educational Measurement: Issues and Practice*, 38-51.

Kihlstrom, J. F., Eich, E., Sandbrand, D., & Tobias, B. A. (2000). Emotion and memory: Implications for self-report. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, & H. S. Kurtzman (Eds.), *The Science of Self-Report: Implications for*

Research and Practice (pp. 81-99). Mahwah, NJ: Lawrence Erlbaum Associates.

Knowles, E. A. (1988). Item context effects on personality scales: Measuring changes the measure. *Journal of Personality and Social Psychology, 55*, 312-320.

Knowles, E. S., & Byers, B. (1996). Reliability shifts in measurement reactivity: Driven by content engagement or self-engagement? *Journal of Personality and Social Psychology, 70*, 5, 1080-1090.

Knowles, E. S., Coker, M. C., Scott, R. A., Cook, D. A., & Neville, J. W. (1996). Measurement-induced improvement in anxiety: Mean shifts with repeated assessment. *Journal of Personality and Social Psychology, 71*, 2, 352-363.

Leary, L. F., & Dorans, N. J. (1985). Implications for altering the context in which test items appear: A historical perspective on an immediate concern. *Review of Educational Research, 55*, 3, 387-413.

Levine, H., Huff, J. W., Wagner, S. H., and Sweeney, D. (1998). The moderating influence of attitude strength on the susceptibility to context effects in attitude surveys. *Journal of Personality and Social Psychology, 75*, 2, 359-373.

Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement, 3*, 102-122.

Linacre, J. M. (1989). *Many-Faceted Rasch Model Measurement.* Chicago, IL: MESA Press.

Lord, F. M. & Novick, M. R. (1968). Measurement in psychology and education. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores.* (pp. 13-24). Reading, MA: Addison-Wesley.

Lunz, M. E., Wright, B. D., & Linacre, J. M. (1990). Measuring the impact of judge severity on examination scores. *Applied Measurement in Education, 3*, 4, 331-345.

Markus, H. (1977). Self-schemata and processing information about the self, *Journal of Personality and Social Psychology, 35*, 2, 63-78.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

Menon, G., & Yorkston, E. A. (2000). The use of memory and contextual cues in the formation of behavioral frequency judgments. In A. A. Stone, J. S., Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, et al. (Eds.), *The science of self-report: Implications for research and practice.* (pp. 63-79.). Mahwah, N.J., US: Lawrence Erlbaum Associates Publishers.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equations of state space calculations by fast computing machines. *Journal of Chemical Physics, 21*, 1087-1091.

Millar, M. G., & Tesser, A. (1986). Thought induced attitude change: The effects of schema structure and commitment. *Journal of Personality and Social Psychology, 51*, 259-269.

Mischel, W. (1968). *Personality Assessment*. New York: Wiley.

Mischel, W., & Shoda, Y. (1999). Integrating dispositions and processing dynamics within a unified theory of personality: The cognitive-affective personality system. In L. A. Pervin & O. P. John (Eds.) (pp. 197-218). *Handbook of Personality: Theory and Research.* New York: Guilford Press.

Mislevy, R. J. (1982, March). *Five steps toward controlling item parameter drift.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*, 2, 159-176.

Muraki, E. & Bock, D. (2002) *PARSCLE 4.1 Computer program.* Chicago: Scientific Software International, Inc.

Nasby, W. (1989). Private self-consciousness, self-awareness, and the reliability of self-reports. *Journal of Personality and Social Psychology, 56*, 6, 950-957.

Neal, R. M. (1997). *Markov chain Monte Carlo methods based on "slicing" the density function.* Technical Report 977, Department. Statistics, University of Toronto.

Nisbett, R. & Ross, R. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, N. J.: Prentice-Hall.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231-259.

Ostrom, T. M., Betz, A. L., & Skowronski, J. J. (1992). Cognitive representation of bipolar survey items. In N. Schwarz & S. Sudman (Eds.), *Order Effects in Social and Psychological Research* (pp. 297-311). New York: Springer-Verlag.

Patz, R J., and Junker, B. W. (1999a). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics, 24*, 146-178.

Patz, R J., and Junker, B. W. (1999b). Applications and extensions of MCMC in IRT: Multiple Item Types, Missing Data, and Rated Responses. *Journal of Educational and Behavioral Statistics, 24*, 4, 342-366.

Paulhus, D. L. (2002). Socially desirable responding: the evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The Role of Constructs in Psychological and Educational Measurement* (pp. 49-69). Mahwah, NJ: Lawrence Erlbaum Associates.

Raftery, A. E., & Lewis, S. M. (1992). One long run with diagnostics: implementation strategies for Markov chain Monte Carlo. *Statistical Science, 7*, 493-497.

Robie, C., Zickar, M. J., & Schmit, M. J. (2001). Measurement equivalence between applicant and incumbent groups: An IRT analysis of personality scales. *Human Performance, 14*, 187-207.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations of p-values in

composite null models. *Journal of the American Statistical Association, 95*, 1143-1172.

Samejima, F. (1969). Estimation of ability using a response pattern of graded scores. *Psychometrika Monograph,* No. 17.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54* (2), 93-105.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linden, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B, 64,* 583-604.

Spiegelhalter, D. J., Thomas, A., Best, N., & Lunn, D. (2003). *WINBUGS Version 1.4 User's Manual [Computer Software manual]*. Cambridge, UK: MRC Biostatistics Unit.

Steinberg, L. (1994). Context and serial-order effects in personality measurement: Limits on the generality of measuring changes the measure. *Journal of Personality and Social Psychology, 66,* 2, 341-349.

Sykes, R. C., & Fitzpatrick, A. R. (1992). The stability of IRT b values. *Journal of Educational Measurement, 29,* 3, 201-211.

Tourangeau, R. (2000). Remembering what happened: Memory errors and survey reports. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, & H. S. Kurtzman (Eds.), *The Science of Self-Report: Implications for Research and Practice* (pp. 29-47). Mahwah, NJ: Lawrence Erlbaum Associates.

Tourangeau, R., & Rasinski, K. (1988). Cognitive processes underlying context effects in attitude measurement, *Psychological Bulletin, 103*, 299-314.

Van der Linden, W. J. & Hambleton, R. K. (1997). Item response theory: Brief history, common models, and extensions. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 1-28). New York: Springer.

Walter, O. B., Becker, J., Bjorner, J. B., Fliege, H., Klapp, B. F., & Rose, M. (2007). Development and evaluation of a computer adaptive test for 'Anxiety' (Anxiety-CAT). *Quality of Life Research, 16*, 143-155.

Wang, W-C., & Liu, C-Y. (2007). Formulation and application of the generalized multilevel facets model. *Educational and Psychological Measurement, 67*, 4, 583-605.

Zickar, M. J., & Ury, K. L. (2002). Developing an interpretation of item parameters for personality items: Content correlates of parameter estimates. *Educational and Psychological Measurement, 62*, 19-31.

Zwick, R. (1991). Effects of item order and context on estimation of NAEP reading proficiency. *Educational Measurement: Issues and Practice, 10*, 10-16.

Vita

Heather Hayes is a Quantitative Psychology Ph.D. Candidate at Georgia Tech and Graduate Research Assistant at the Georgia Tech Research Institute. Heather has received her M.S. in Industrial-Organizational Psychology from Virginia Tech. Her research interests center on psychometrics, Item Response Theory (computer adaptive testing and automatic item generation), latent mixture modeling, and cognitive modeling of response processes in various types of trait measures, particularly personality.